

European principles worth sustaining

As the time approaches for negotiations over the next Framework programme of research and development, Europe's governments have begun to make their views known. Some aspects of the current programme may need protection.

BRUSSELS paperwork tends to make eyes glaze over and pulse-rates sink towards frequencies that would set alarms beeping in hospital theatres. People seeking funds from the European Commission might be forgiven for suspecting that their applications induce such reactions in administrators, so snail-like is the process. But although there is chronic concern about the forests consumed and the time spent by researchers and officials in the commission's research and development Framework programmes, there are other more fundamental issues that can be addressed only when a new Framework is being negotiated. That season is now upon us, as governments and organizations present their initial positions (see page 634) as to how the Fifth Framework, due to start in 1998, should compare with the Fourth (1994–98).

Since the inception of the European Union's research and development (R&D) programmes, Europe's industrial prowess has been their primary motivator. "Cohesion" — the desire that programmes should in particular boost R&D and industries in less technologically developed member states — has also been high on the agenda. Should those driving forces continue to be so dominant?

Certainly there continue to be reasons to worry about Europe's commercial competitiveness. A green paper (discussion document) issued by the commission at the end of last year highlighted the continuing weakness of Europe in innovation compared to the United States and Japan. But there are many contributing factors. R&D collaboration programmes, however successful, have difficulty in paying off if they are carried out by companies competing against those in environments more favourable for business investment, where higher priority is given to R&D and where the business culture is more entrepreneurial (the United States) or steadfast (Japan).

Nevertheless, earlier Framework programmes such as ESPRIT (information technologies) and RACE (advanced telecommunications) have led to some identifiably commercial spin-off. Although the programmes were supposedly supporting work far upstream from the market, some returns came promptly — after all, it is often hard to tell whether a particular piece of research will turn out to be far from or near to the delivery of a valuable product enhancement. But industrialists will point to a less tangible outcome as a prime benefit that would otherwise not have arisen: contacts with others across Europe, whether in competing companies or positioned somewhere useful in the supplier/customer chain. That has surely led not least to an enhanced exploitation of the European market. Furthermore, the company politics has a lot going for it: industrial researchers find it easier to get internal funds for research if they can tell their business managers that European competitors are also doing it, and if they can tell otherwise tightfisted finance directors that the commission will match it.

In short, the industrial R&D programmes have by and large played a useful function without distorting research priorities, and deserve continued strong support. In the current Framework, such programmes have made a virtue of including service organizations representing end-users of the technologies — banks, for example. The particular benefits of that remain to be

demonstrated, especially if such involvement has led to subsidy of near-market product development that should have been supported by private investment. But the idea explicit in the UK government's proposals that businesses, including such end-users, should have a greater say in developing the principal thrusts of the Fifth Framework has a lot to be said for it, and should strengthen the development of commercial innovation that the green paper says is inadequate.

The main beneficiaries of the industrial technology programmes have been larger companies, not least because they have the resources to cope with the paper-pushing. To its credit, the commission has established schemes to encourage smaller companies to get involved. Regional agents have been appointed to encourage small and medium enterprises (SMEs) to apply for a small 'exploratory' grant that will in turn give them the scope to develop a full-blown proposal for involvement in a programme. Finding partners is the biggest obstacle, despite the growing use of databases. But without partners, small organizations tend to lack the resources required to exploit the results of their project. For that reason such stimulation is valuable, but audits of its success are also required.

What of cohesion? Support for SMEs is an important factor in its pursuit, given the paucity of corporate giants in less developed parts of the European Union. And everyone understands another significant lever: the unofficial fact that Framework grants are preferentially awarded to those who include partners from less developed regions. But the new government proposals for the Fifth Framework appear to signal a weakening of support for cohesion. France's research minister, François d'Aubert, last week declared flatly that too much money is being spent on it. The United Kingdom is notably silent in its support for the idea, and implicitly negative in its prominent recommendation that only the best research should be supported.

Support for cohesion should, on the contrary, be maintained. Companies in less developed regions often have to battle against weak local infrastructure and higher interest rates. They, as well as university groups hindered by regional obstacles, can gain disproportionately through contact with those at the leading edge. Who else but the commission will readily issue cash for travel and meetings?

The Training and Mobility of Researchers (TMR) programme is unique within the Fourth Framework for its emphasis on support for fundamental research. Some wish to enhance its industrial relevance. Unless additional money is forthcoming for the purpose, that too should be opposed. Despite the richness and fertility of Europe's scientific landscape, there are many obstacles that discourage young scientists from crossing national boundaries. That and the fact that there is too little support elsewhere for such mobility makes it all the more important for the TMR programme's activities in fundamental research to be sustained, not only to help the regionally disadvantaged but also to make the most of Europe's very best. That principle is surely what makes Framework paperwork worth the effort. □



NIH bucks political trend to win increased funds from Congress

Washington. A subcommittee of the US House of Representatives last week approved a generous increase in funding for the National Institutes of Health (NIH) next year. If eventually approved by Congress, this would result in an increase of almost seven per cent for the biomedical agency, at a time when dozens of other government programmes are having their funds cut.

The proposed increase comes from the subcommittee on labour, health and human services and education of the House Appropriations Committee, which early last Friday (14 June) passed a spending bill for the fiscal year 1997, raising NIH funding by \$820 million — or 6.9 per cent — from \$11.9 billion in 1996 to \$12.7 billion in 1997.

In a surprise move, the Republican-dominated subcommittee also removed a provision extending a ban on government funding of human embryo research.

The bill includes \$90 million for a new 250-bed Clinical Research Center (CRC) planned for the NIH campus, considerably less than the \$310 million requested by the Clinton administration. But it increases the rest of NIH funding, most of which is spent on research, by 6.5 per cent, \$765 million above the 1996 level.

The increase is allocated fairly evenly across the NIH institutes. The two largest — the National Cancer Institute and the National Heart, Lung and Blood Institute — would both receive increases of 6.1 per cent, while the National Institute of Allergy and Infectious Diseases would see its budget grow by 7.5 per cent, well above the anticipated rate of inflation.

Harold Varmus, the director of NIH, was cautious in welcoming the subcommittee's decision to increase the research budget. "It's a pleasing number, but it's only the first step in the process," he said on Monday.

In contrast, Ralph Bradshaw, president of the Federation of American Societies for Experimental Biology, which has lobbied hard for increased funding for biomedical research, said his organizations were "really very, very pleased". The proposed 6.5 per cent increase would make things "as rosy as they have been at the NIH for some time".

Few other agencies in the \$65.7-billion bill passed by the subcommittee, which funds three major government departments, benefited from the largesse showed towards the NIH. This prompted complaints from liberal Democrats that other programmes, such as education, had suffered as a result.

But John Porter (Republican, Illinois), chairman of the subcommittee, defended

NIH's allocation. "It's something only government can do. There's no profit motive in basic research," he said. "It pays for itself in terms of health care cost savings thousands of times over. It's perhaps the most efficient spending government ever does."

The bill will become law only after it has been approved by the full House Appropriations Committee, the full House, parallel Senate bodies and President Bill Clinton. But the NIH funding increase, largely the work of Porter, is expected to set the mould for House legislators.



Porter: praised cost-effective work by NIH.

The reversal of the ban on NIH funding for human embryo research was passed by nine votes to four, with Porter and three other Republicans joining Democrats on the subcommittee to end the ban.

Porter cited recent testimony by Eric Wieschaus of Princeton University, a co-winner of last year's Nobel prize for physiology or medicine, in arguing for the ban's reversal.

If the reversal is maintained, the ban would be lifted from the beginning of the fiscal year on 1 October. But the author of the current ban, Jay Dickey (Republican, Arkansas), vowed to reinsert it in the bill at the full Appropriations Committee.

In a separate vote, the subcommittee defeated a bid by Democrats to give back to

the Office of AIDS Research (OAR) full control over the money spent by NIH on AIDS research, \$1.4 billion in the current fiscal year. The Democratic attempt, if successful, would have given the OAR the power to distribute directly the huge AIDS research budget to the NIH's 24 institutes, centres and divisions.

This power is mandated by a 1993 law authorizing the NIH budget, but is not currently in place, as Republicans circumvented it in a 1996 spending bill. The effort to restore a 'consolidated' budget to the OAR, introduced as an amendment by Nancy Pelosi (Democrat, California), was defeated by nine votes to five, the subcommittee splitting along party lines.

After the vote, Mark Harrington of the Treatment Action Group, an AIDS advocacy group, complained that the question of the OAR's authority had become "a political football". He pointed out that in March, a 118-member expert panel chaired by Arnold Levine of Princeton University had recommended that the OAR retain its fund-distributing duties.

Other AIDS activists said that implementation of Levine's report on restructuring AIDS research at NIH would be badly damaged if the subcommittee's decision stands. "Without budget authority [for OAR] the Levine committee report is waste paper," said Gary Rose, of the AIDS Action Council, which is based in Washington, DC.

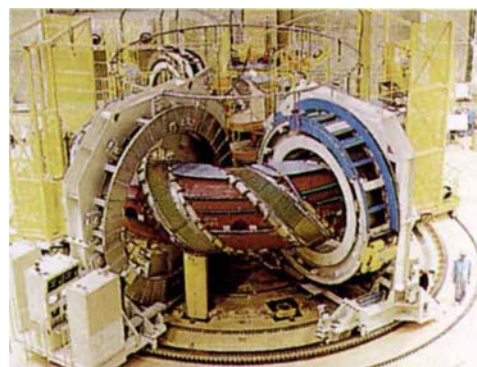
But Levine himself disputed that. "I don't believe that the implementation ... of our report [depends] on a consolidated budget," he says.

Meredith Wadman

Japan puts a twist in fusion efforts

Tokyo. The world's largest helical fusion device (right), being built by Japan's National Institute for Fusion Science (see *Nature* 335, 104; 1988), took a major step towards completion last month with the winding of the superconducting coil from 36 km of superconducting cable.

The 250-tonne coil will be installed by the end of this year, and the device is expected to commence operation in February 1998. Scheduled to cost about ¥50 billion (US\$470 million), it is under construction at Toki in Gifu Prefecture, in central Japan. The helical fusion



device will generate a steady-state plasma at a near-critical temperature of 100 million °C.

Stephen Barker

Germany seeks limited European projects...

Munich. Germany is expected to propose that, for the first time, some specific research projects funded under the European Commission's next four-year Framework programme, which starts in 1998, should be restricted to a limited number of member states.

The proposal is likely to be included in a position paper that the German government is due to publish later this month in response to the commission's request for formal input from member states before it puts forward its own proposal for the next Framework programme by the end of the year.

Like those of Britain, the German suggestions emphasize the need to improve efficiency while controlling application numbers. For example, it would like to see the management of some programmes delegated to organizations with appropriate expertise within the member states.

This has already been tried out with a plant biotechnology project called AMICA, which has been handled by the John Innes Research Institute in the United Kingdom. But the commission has been reluctant to

European governments last week presented to the European Commission in Brussels their recommendations for the Fifth Framework programme, its four-year, ECU12.3-billion (US\$15.1-billion) programme of joint research, which is due to commence in 1998. We report here proposals from Germany, France and Britain. Submissions from other countries, as well as from bodies such as the European Science Foundation, will be covered in future weeks. The commission's final proposals are due by the end of the year.

experiment further with this idea because it is not clear that it has a legal right to do so.

In order to limit the number of applications, Germany would like to introduce a concept of 'variable geometry', under which participation in certain programmes would be restricted to a few identified countries with appropriate research strengths, rather than all 15 member states.

One such programme might be in aero-

nautics. This position has the support of at least the Netherlands, and there is provision for such a concept within the Maastricht Treaty. But countries likely to be excluded are unlikely to be enthusiastic.

The German paper points to the need for mechanisms to recognize and correct problems as they arise during the running of the programme. It also argues that, in defining research objectives, more use should be made of technology foresight programmes.

At the same time, Germany would like the number of programmes to be reduced considerably, both to reduce the number of applications and to concentrate funds. It suggests seven: informatics, production processes and material sciences, environmental and marine sciences, life sciences, energy, transport and mobility, and socioeconomic research.

Germany specifies themes that it believes should cut across its suggested programmes. These are ageing, robotics, vaccines and viral diseases, clean power stations, telemedicine, and the car of the future — subjects of existing or possible future task forces.

It strongly supports the expansion of the task force concept introduced last year by the research commissioner, Edith Cresson, as a means of coordinating research on specific topics across programmes and across commissions.

It also supports the move by the commission to use proposed additional funds promised to the Fourth Framework programme, which have yet to be approved by the Council of Ministers, to test the effectiveness of at least some of Cresson's task forces. But it criticizes the 'top-down' process by which task forces have so far been chosen.

Germany is particularly keen for the increased development of research programmes geared to cooperation with countries of Central and Eastern Europe

Alison Abbott

...as France backs competitiveness

London. France has claimed that the costs of running the Framework Programme are "too expensive", and has suggested that "bold and imaginative" solutions are needed to remedy a situation that "could challenge the very existence of such an investment by the European Union".

It has also proposed, in common with its main European partners, that the main thrust of the Fifth Framework programme should be more clearly focused on increasing the competitiveness of industry.

At the same time, it says there should be a specific programme of basic research linked closely to technology-based programmes, based on concern that an excessively utilitarian approach could undermine the long-term investments needed to maintain 'upstream' research.

France would also like to see a programme to support 'technological leaps', organized through what it calls a 'bottom-up' procedure, allowing new sectors to be covered and assuring flexibility and adaptability in the programme.

These and other proposals were put forward last week by François d'Aubert, the secretary of state for research, following consultation with industry, government and academic groups in France.

Speaking in Paris at a meeting held to discuss the results of this consultation exercise, d'Aubert said that analysis of the responses had revealed a "remarkable convergence" in the comments received, and also that informal contacts with

British and German officials "tie in with the main thrust of our analysis".

According to d'Aubert, the current Framework programme suffers from two structural defects, both of which require correction. The first is that it spends too much money on achieving the European Union (EU)'s goal of "cohesion" — supporting research of limited strategic value in "less competitive" member states.

A second defect is the slowness and cost of administering the programme, including the costs incurred by applicants which, given the low success rates in some programmes, could, in aggregate, exceed the value of the grants.

David Dickson

...and Britain pursues key role for 'users'

Munich. Britain has proposed that the joint research efforts of the European Union (EU) should be remodelled along similar lines to those on which its own science and technology policies have been based since the publication of its domestic white paper (policy document) on science and technology in May 1993.

In particular, Britain says that a far greater role should be given to the 'user community', primarily the industrial sector but also government departments and other public institutions, in setting the research agenda for the Fifth Framework programme, which starts in 1998, to ensure that the programme's output is "genuinely responsive to Europe's needs".

In a position statement published in London last week, the UK government also recommends a large increase in funding for activities relating to the dissemination and exploitation of research results, increasing their share of the Framework budget from one to five per cent.

Other suggestions are increased efforts to monitor and evaluate research projects, improvements to the grant application procedure — including "clearer, more consistent and more timely publicity and advice to applicants" — and contracting out some aspects of contract management.

But, not surprisingly in view of its history, Britain opposes any increase in the programme's budget; it says that the over- ▶

► all budget for the new Framework programme should not be greater than that of its predecessor, the Fourth Framework, which started in 1994.

The main thrust of the British position is that Framework should be organized around “objective driven themes” that address the “medium-term technological needs of European industry” and the need to “improve the quality of life” — both important motifs in the 1993 white paper — as well as support for European policies more broadly.

To achieve the first two goals, it suggests setting up a number of advisory groups, made up mainly of individuals representing user communities who, acting as ‘proxy customers’, would make recommendations on which the Commission would establish its strategic priorities.

A limited number of research and development programmes would then be identified as contributing to each of these priorities. These would then be aggregated to form “technology-specific delivery programmes”, each allocated an agreed budget and monitored by what are described as “regulatory committees”.

Britain suggests a number of possibilities for the strategic priorities, ranging from “sustainable farming and fishing” to “tomorrow’s car”. Six proposed programmes would contribute in varying degrees to the achieve-



European synchrotron research: showing how collaboration can be effective.

ment of these goals, from information and communication technologies to life science and medical technologies.

Officials admit that this strategy has some similarities to the task force approach currently being pursued by the commission. But they claim it also has important differences. “Task forces are a bit too ‘top-down’,” says Ian Taylor, Britain’s minister for science. “They make a good starting point; but a task force that seems to anticipate [the outcome of a research programme] is not the most effective way of doing things.”

At the same time, Taylor is cautious about Germany’s suggestion of introducing a “variable geometry” for new programmes (see above left). “This is an untested idea. We would not close our eyes to it but would want to be convinced that it would have no budgetary implications, and want to see much more of what the idea would mean in practice.”

A. A. & D. D.

US bid to eliminate gene therapy panel under fire

Washington. Opposition is growing to a proposal by Harold Varmus, the director of the US National Institutes of Health (NIH), to disband the biomedical agency’s Recombinant DNA Advisory Committee (RAC).

The 25-member panel of scientists, ethicists, lawyers and others is responsible for reviewing and approving novel gene-therapy protocols. Varmus wants to hand over this function to the Food and Drug Administration (FDA), retaining at the NIH a committee that would convene conferences on novel gene-therapy topics.

But an official announcement of the proposal, due to have been published in the *Federal Register* several weeks ago, has been delayed, as opposition to the proposed dissolution of the committee has grown. Last week, for example, Lana Skirboll, associate director for science policy at the NIH, together with other NIH officials, briefed Senate aides at a meeting reported by one participant to have included “heated exchanges”.

Individual senators who have voiced concern include David Pryor (Democrat, Arkansas), who is preparing a letter to Varmus for which he plans to gather the signatures of other members of the Senate and House of Representatives. “NIH should be trying to revise RAC, not abolish it,” Pryor said in a statement, pointing out that the Congress relies on the panel for a

“public dialogue” on issues raised by recombinant DNA and gene therapy.

Patient advocates also argue that the closed-door reviews conducted by the FDA — which is not charged with deciding the ethical merit of proposed experiments but only their safety and efficacy — are a poor substitute for public RAC meetings.

Skirboll says that the FDA takes ethical questions seriously, and would voluntarily refer to the NIH under the new system. This, she says, will be more forward-looking, anticipating controversial developments, such as the potential use of HIV as a vector, and convening conferences to outline recommendations well before protocols are developed.

One of those who approve the proposal is Inder Verma, a professor of molecular biology at the Salk Institute, in La Jolla, California, and chairman of a review committee that last year recommended that the RAC continue to approve protocols that raise new issues. He describes the proposal as “very sensible”.

Verma says that a new standing advisory committee to NIH’s Office of Recombinant DNA Activities (ORDA), proposed by Varmus, will “essentially be doing exactly the same thing” that the RAC now does, by holding public conferences on novel and controversial gene-therapy issues.

Meredith Wadman

UK scheme aims to make biotech pay

London. The British government this week unveiled a series of initiatives designed to boost the country’s emerging biotechnology industry, and in particular to encourage more laboratory scientists to shake off their perceived inhibitions about entering the world of entrepreneurs.

The ‘Crusade for Biotechnology’, which spans several government departments, was launched in London by Ian Lang, President of the Board of Trade. It will provide scientists with management advice, business contacts and access to new money for ideas that can turn biotechnology research into commercial success.

Encouraged by a recent report from the accountancy firm Ernst and Young that places Britain at the top of Europe’s biotechnology league table — the turnover of its bioscience companies was £310 million (US\$465 million) in 1995, second only to the United States — the government is keen for the country to maintain this position. The figure accounts for one-third of biotechnology revenues in Europe.

“The race is only just beginning,” said

Lang. “We cannot afford the luxury of self-congratulation or, worse, the folly of complacency,” he added.

The centrepiece of the new initiative is a £25-million fund providing access to capital for new biotechnology companies able to exploit technology developed at laboratories run by the Medical Research Council (MRC). The fund will last five years and be operated by a company owned by the MRC but governed by an independent board.

The initiative also includes £7 million for schemes offering advice to scientists on intellectual property rights, professional guidance on setting up pilot businesses — known as ‘incubators’ — able to demonstrate the commercial potential of an idea, and three £50,000 prizes to encourage business ideas from young researchers.

Meanwhile, the Biotechnology and Biological Sciences Research Council is to collaborate with the Royal Society and the Engineering and Physical Sciences Research Council in a new scheme promoting contact between industry and university-based scientists.

Ehsan Masood

Legal restrictions block NASA 'institutes'

Washington. A year-long effort to establish privatized 'science institutes' at several US space agency field centres appears to have foundered because of legal restrictions on government employees transferring to the private sector.

The National Aeronautics and Space Administration (NASA) had asked for language to be included in its 1997 authorization bill lifting post-employment restrictions on government scientists who might want to convert to contractor status under the institute plan — particularly in relation to their ability to work on contracts they had once managed, and to keep their retirement pay and other benefits.

But the Office of Government Ethics and the Office of Personnel Management both opposed exempting NASA from rules applied to other agencies, and the language was deleted from the authorization bill. As a result, Daniel Goldin, the agency's administrator, has told centre directors in a letter that "we will discontinue our efforts to establish science institutes as previously planned at the Ames Research Center [in California] and the Lewis Research Center [in Cleveland, Ohio]." NASA science managers had maintained that, without 'legislative relief', the science institutes plan might not be viable.

An institute for space biomedical research at the Johnson Space Center in Texas will, however, continue as planned, as

it is not expected to draw heavily on government scientists for staff. NASA will be seeking partners for the biomedical institute later this summer. Other proposed institutes at the agency's Langley and Marshall centres were effectively shelved last year because of a lack of clear definition.

Goldin's decision appears to end the troubled saga of the science institutes, which suffered from a vagueness of purpose and a lack of interest among potential university sponsors (see *Nature* 380, 7; 1996). France Cordova, NASA's chief scientist, says that any further efforts to strengthen the agency's ties with the outside research community will now fall to each individual centre. In his letter to centre directors, Goldin said only that "[w]e will consider other center options that will continue to enhance the quality of the science at the respective field installations".

Exactly what those options are is not clear. William Berry, head of Ames' space directorate, says there is some relief among staff scientists now that uncertainty about the fate of the institutes is over. Rather than create one large umbrella institute, he says, Ames will now work with local universities to set up "new relationships" in specific disciplines, and perhaps consider creating several small institutes.

Berry says that one university has already shown interest in establishing a new astrophysics department, which could collaborate

regularly with Ames staff. The Ames science directorate plans a conference this autumn to stimulate additional relationships with outside researchers.

In the long run, Berry expects work on the space station and on NASA's proposed 'origins' programme — which covers disciplines ranging from cosmology to the search for extraterrestrial life — to provide new lines of business for space researchers. But the space directorate faces a funding cut of 20 per cent in the next several years, which suggests that some of the current Ames work force will have to go.

One research task that might be "outsourced", says Berry, involves the large animal centrifuge designed for life sciences research on the space station. He says that he can envisage centrifuge research being devolved to a privatized institute similar to — but much smaller than — the Space Telescope Science Institute in Baltimore, Maryland.

But even if such mini-institutes are established, current conflict-of-interest laws will prevent government scientists from working there. The same restrictions will apply to the Lewis centre's microgravity research programme, which was another likely candidate for privatization. As a result, NASA staff scientists are now facing the same uncertain future that they faced more than a year ago, when the institutes were first proposed.

Tony Reichhardt

Rapid response team urged for space station research

Washington. The research plans of the US National Aeronautics and Space Administration (NASA) and its partners for the international space station neglect engineering and technology research that could help reduce its operational costs and improve its performance, according to the National Research Council (NRC).

But a report published last week by the NRC, the operating arm of the National Academy of Sciences, says that it is not too late to add more experiments to the line-up — provided that NASA quickly forms a 'rapid response' team to suggest modifications to the station before it begins construction next year.

The NRC panel was chaired by David Bodde of the Midwest Research Institute in Kansas City, Missouri. It says that the station could still be a valuable platform for in-space studies of electric-power generation, robotics, propulsion, life support and several other areas of technology — not least of which is the behaviour of large rigid structures in space.

But as plans for the station have been scaled back in recent years, many such experiments have been dropped. The report

Testing: time for NASA to add experiments?

lists nearly a dozen 'disappearing technologies', including advanced 'solar dynamic' power generation and closed-loop environmental-control systems that would recycle water and oxygen. Many of these had been studied in detail before NASA dropped them from the plan, and the panel believes they could be put back in.

Other suggestions are that NASA should develop a 'road map' for engineering and technology research, ensure that experiment

NASA interfaces on the station are similar to those in laboratories on the ground, and investigate the feasibility of placing instruments on the station to help monitor its structural dynamics and the space environment.

NASA should also involve industry and academic institutions in selecting engineering and technology experiments, says the panel. One of the most controversial suggestions — and probably the least likely to be followed — is that the space agency should auction off perhaps 15 per cent of station resources to commercial researchers, rather than try to decide itself what work has commercial potential.

Many of the criticisms — such as the lack of common interfaces and the difficulty in proposing technology experiments — have been raised by other panels in the past. Indeed, some committee members say NASA is following its usual tendency of focusing on getting new hardware into space rather than on how it will be used.

Nevertheless, Daniel Goldin, the administrator of the agency, was enthusiastic about the report when briefed on its proposals last week, and may soon appoint a task force to look into its recommendations.

T. R.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Governments urged to back Internet use

Snekkersten, Denmark. Political leaders from 27 industrialized countries last week took the first steps towards greater international collaboration to encourage and regulate the growth of the use of the Internet for research.

The move came at a conference on the "Global Research Village", organized by the Organization for Economic Cooperation and Development (OECD), and chaired by Frank Jensen, Denmark's minister of research and information technology.

Much of the two-day meeting — which was held near Helsingør, the Elsinore of Shakespeare's *Hamlet* — was spent debating the role that governments could and should play in developing research networks. "Governments haven't yet defined their role," argued Michael Osborne, the OECD's deputy director for science technology and industry, at the opening session.

Two broad areas of consensus emerged. The first was the need for governments to ensure that the research community has an adequate networking infrastructure by resolving the current congestion on the Internet and lowering research networks' operating costs (see *Nature* 380, 377; 1996).

The second was that international agreements harmonizing legal, regulatory and commercial practices are needed to ensure that the Internet remains an open system able to contribute to economic growth and social development in all countries, including those in the developing world.

Jensen said that research networks could not rely on commercial expansion and improvement of the Internet. Public investment is needed to give researchers the infrastructure they need, he said, adding that governments should facilitate the development of high-speed research networks.

But the appropriate form of government intervention was hotly debated at the meeting. Robert Cailliau, director of the World Wide Web at the European Laboratory of Particle Physics (CERN) argued that, while government researchers still need funding to pay for their access to the Internet, financing of the extra bandwidth needed on the network will not come from government but from expansion of commercial services.

Other delegates argued, however, that governments could usefully provide funding to pump-prime new developments, such as the US National Science Foundation's support for new high-speed backbones (see *Nature* 380, 93; 1996). Furthermore, one delegate said, cooperation between OECD governments on providing adequate links between existing networks is insufficient.

Several delegations also insisted that liberalization of telecommunications is a prerequisite of low-cost research networks. J. M. M. Ritzen, the Dutch minister of education, culture and science, pointed out

that Europe's telephone companies charge customers ten times as much for a new service as do their US counterparts.

An even bigger hindrance, he said, is the reluctance of telephone monopolies to meet demand for high-speed connections. OECD governments have "an important part to play in remedying the inadequacies of the telecoms infrastructure that stand in the way of scientific communication", said Ritzen.

José Mariano Gago, Portugal's minister for science and technology, went further, suggesting that research organizations should be required to pay only a fraction of



the real operating costs of networks. Jensen said governments have the power to negotiate better conditions from telecommunications operators, pointing out that he has obtained a 65 per cent reduction in the costs of research networks from Danish telecoms.

Another area where OECD governments could usefully cooperate, participants said, is in developing international "legal, regulatory, and commercial regimes". They agreed, for example, that an effective legal framework is needed within existing intellectual property regimes to cover issues raised in the transfer of scientific information between the public and private sector.

One result of the lack of international agreement on legal issues is that, under

contracts from some US telecommunications companies, European research networks could, in principle, be sued if any of their researchers use their US links to access pornographic sites.

Indeed, Cailliau said that while he had originally opposed greater government control of the Internet, he now feels this was needed to bring greater order. Issues such as preventing criminal use of the Web require internationally binding agreements, he said. "A happy village is not on the cards unless politicians make it happen," said Pat Rabbitte, Ireland's minister of commerce, science and technology.

The meeting also strongly recommended that OECD governments support the development of virtual electronic research libraries and the digitization of research documents. Such libraries, it said, should be open to all researchers worldwide, both in the public and private sectors.

To promote greater collaboration between industry and academic institutions, the meeting also said that OECD countries should explore the setting up of international "virtual research centres of excellence" and the use of networks to provide better access to expensive science facilities.

Meanwhile, many delegates expressed concern that the flexibility offered by electronic publishing risks reducing the quality of scientific information, and with it the reputation of the scientific community. The meeting unanimously supported the need to maintain high standards of quality assurance in electronic publishing.

But while agreeing that peer review should be the principal means of doing so, it also said that governments should "explore" the need to subsidize the setting up of 'e-print' archives in all disciplines.

Although the 'Elsinore agenda' agreed at the end of the meeting merely represents its conclusions and does not imply any commitment on the part of the governments represented, Jensen says it is a first step towards putting research and the Internet on the political agenda.

Declan Butler

Germany gears up for high-speed network

Munich. After years of lagging behind its neighbours, Germany is about to become the second country in Europe to offer its scientists a national, very-high-speed broadband data network.

Next month, data transmission on the newly established national broadband scientific network, called B-Win, which links around 50 university departments and research institutes, will be increased in speed from 34 megabits to 155 megabits per second.

In the past, delays in modernizing

communications systems have caused Germany many problems. In 1993, for example, it lost a bid to host the European Bioinformatics Institute, an outstation of the Heidelberg-based European Molecular Biology Laboratory, partly because of its poor electronic datalinks (see *Nature* 361, 383; 1993).

B-Win is run by the Deutsche Forschungsnetz, an organization based in Berlin and similar to UKERNA, a British body which operates the UK academic network JANET. **Quirin Schiermeier**

US panel backs new approach to risk

Washington. A new approach to assessing environmental and public-health risks, allowing the public and scientists to become involved in discussing risks before they are given a formal assessment, has been suggested by the National Research Council (NRC), the operating arm of the US National Academy of Science.

According to officials of leading federal agencies, the report, *Understanding Risks: Informing Decisions in a Democratic Society*, published in Washington DC last week, could have far-reaching implications for the way in which government uses science to reach decisions about managing such risks.

The suggested approach immediately came under fire from industrial groups, which claim that excessive participation by 'stakeholders' — including industry, environmentalists and the general public — could swamp objective, scientific risk-assessment.

But Harvey Fineberg, dean of the School of Public Health at Harvard University, and chair of the panel that produced the report, argues that government stands to make both better and "more efficient" decisions by adopting the recommended procedures.

The proposals are seen by some of the NRC panel members as a significant departure from existing practice, which is largely based on a 1983 NRC report, *Risk Assessment in the Federal Government*, widely referred to as the 'Red Book' after the hue of its cover. "Maybe it should be called the 'misread book'," says Warner North of Decision Focus Inc., Mountain View, Cali-

fornia, and a member of Fineberg's panel.

The 1983 report contained a diagram that has been very widely interpreted as presenting risk management as a linear process, in which scientists research a problem and then formally assess the risk the problem poses to public health.

Agencies are then supposed to consult all interested parties, and to consider social, political and other factors, before introducing regulations. The new NRC approach (see diagram, below) suggests that the consultation and deliberation should take place at the outset.

Officials say that some federal agencies already use the recommended approach, which engages scientists and managers in public discussion from the outset. When

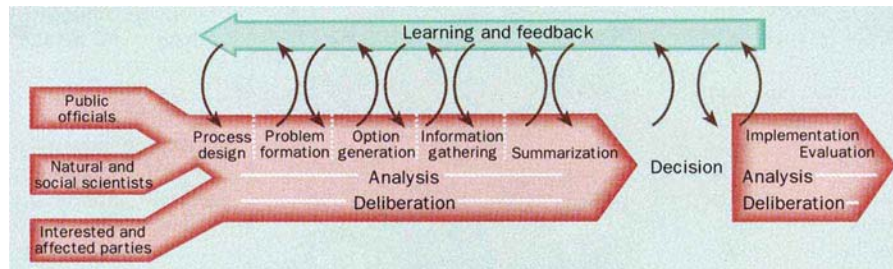
— and that is disquieting to me," he says. Lewis predicts that science, if mixed up with the policy, "would bring an even more confusing dog's breakfast to the table" than it does already. He warns against "an arranged marriage" between science and policy, from which "there will be no divorce".

Some government officials are also concerned about the degree of integration proposed by the panel. Alwynelle Ahl, director of risk assessment at the Department of Agriculture, says risk assessment is a new function in the department, only starting to establish its independence in the agency.

Risk assessors, she says, are wary of "managers coming in and saying 'this is the result I want and you're going to get it'". Ahl endorses the general message of the report.

But she warns that "the challenge will be to test the theory of this report in the cauldron of agency decision-making".

Industry fears that endless debate with the public will cloud clear-cut scientific assessment of risks, especially celebrated



Risky business? The new assessment model involves the public from the start.

military and nuclear sites are cleaned up, for example, the Departments of Defense and of Energy have tried to consult widely before the specific problems of a particular site are assessed, rather than bringing in the public after this has been done to discuss proposed solutions.

But in many other cases, researchers are asked to assess a particular problem — such as the cancer risk from a certain chemical — and then step aside while agencies, industry and public interest groups argue about what to do about it.

"Many scientists have felt that they were [just] being asked for a number," says Lynn Goldman, assistant administrator for pesticides and toxic substances at the Environmental Protection Agency, warmly welcoming the report.

Goldman says that the previous approach does not make full use of scientists' skills, and that they could contribute more through "a back and forth process" with agency managers and the public. "You need an iterative process, involving many people," says North.

But the American Industrial Health Council, an industry group which, with several government agencies, helped to fund the study, has denounced its main findings. Steven Lewis, a senior scientist at Exxon, the oil company, and chair of the NRC's science policy committee, says that science has never had a strong enough voice in risk management.

"In order to correct that, the report proposes to partner science and policy up close

ones such as the cancer risk posed by pesticides, which it considers to be small.

It supports legislation that would require such assessments both to be published and to be used as the main basis for government regulation. Legislation towards this goal was passed by the House of Representatives last year but is now stalled in the Senate.

According to one congressional staff member involved in drafting that legislation, the NRC report is "full of 'kumbaya' stuff" and "doesn't get to the heart of the problem". North says that the proposed structure "should not make scientists more restricted, it should liberate them, and enable them to participate more fully" in defining and solving environmental and public-health problems.

But the report warns that the proposed consultative processes will be wrecked if one party bales out to attack it in court, or by a direct appeal to the Congress to change laws. "This possibility casts a shadow over the entire process," it observes, since it "can increase alienation and mistrust" among those who were reluctant to get involved in the first place.

Coincidentally, a presidential commission on risk assessment and risk management, established under the 1990 Clean Air Act, issued a draft report last week. It criticized what it termed the "highly fragmented, adversarial system" for risk management in the United States, and called for more and earlier public participation in the process.

Colin Macilwain

US-French pact allows N-test collaboration

Washington. The United States and France have signed a wide-ranging agreement to collaborate on nuclear weapons research and to share weapons test data, according to a report published this week in the *Washington Post*.

The agreement, signed on 4 June, is said to give France access to some US test data, as well as the results of computer simulations. It will also allow weapons scientists from both countries to use huge laser facilities that each is planning to build: the French laser is expected to come into operation first.

Both countries maintain the goal of the proposed research is to ensure the safety and reliability of their existing weapons, and not to design new ones.

C. M.

Hostile reception to US misconduct report

Washington. An internal panel this week advised Donna Shalala, the Secretary of the US Department of Health and Human Services (DHSS), to adopt most of the recommendations of a controversial report on scientific misconduct, over the protestations of dozens of scientific groups, including, most recently, members of the advisory board to the National Institutes of Health (NIH) director, Harold Varmus.

The DHHS panel was charged with recommending to Shalala how to implement the report of the Commission on Research Integrity (CRI). The report was mandated by a 1993 law and released last November by the CRI, a 12-member panel of nongovernment scientists, ethicists and lawyers. If adopted, the recommendations would apply to thousands of scientists receiving funds from NIH and other government agencies.

In presenting his panel's 35-page assessment of the CRI report to Varmus's advisory board last Monday, 17 June, William Raub, the head of the panel, said the report "has much to commend it". Many of its recommendations, he said, "would have a positive effect" on research integrity.

The panel's recommendations have been sent to Shalala, who has final authority over whether and how to adopt them. It is not known when she will announce a decision.

The chairman of CRI, Kenneth Ryan, an emeritus professor of obstetrics, gynaecology and reproductive biology at Harvard Medical School, called the panel's product "a fair hearing" and "an appropriate way to go".

The Raub panel deferred judgement on three of the CRI's 33 recommendations — two of them the most controversial. Of the latter, the first concerns definitions. The current law governing NIH and many other government grant recipients defines scientific misconduct as "fabrication, falsification, plagiarism" or other practices that "seriously deviate" from scientific community norms. The National Academy of Sciences (NAS) uses a similar definition. The CRI replaces these words with "misappropriation, interference and misrepresentation", a definition that critics have called vague and too broad.

The Raub panel said that, because of the "decidedly negative tenor" of reaction to the proposed definition, Shalala should make an official request for comment on the definition before making a decision. This request, however, should be delayed to coordinate with the timetable of a working group of the National Science and Technology Council, which is developing a government-wide definition of scientific misconduct.

The other controversial recommendation is a 'whistle-blower's bill of rights'. The Raub panel says that the CRI report, while it evidently "envisioned a judicious balance", creates "a strong initial impression that the Commission was more attentive to rights of

whistle-blowers and the responsibilities of other parties" than vice versa.

Nonetheless, it says, regulations protecting whistle-blowers must, under the 1993 law, be established by Shalala, and the CRI's report should be taken into consideration by the NIH office now charged with developing them, the Office of Research Integrity.

At last Monday's meeting, several members of Varmus's advisory board protested. "This [CRI] report is a disaster" and would create an "extremely destructive" atmosphere, said Marc Kirschner, chair of the department of cell biology at Harvard Medical School, who also complained that the CRI contained no "distinguished" scientists.

Paul Marks, president of Memorial Sloan-Kettering Cancer Center, insisted that watchfulness over integrity "has to be at the institutional level" and cannot be imposed by government. The president of NAS, Bruce Alberts, who also spoke to the NIH group, said the CRI report would be "very disruptive" if adopted, in part because of the

intrusive federal bureaucracy involved.

Ryan, CRI chairman, retorted later that he has heard these arguments before, and that the fact that little has changed since a 1992 NAS report on scientific misconduct justifies his committee's suggestions.

The Raub panel rejected four of the CRI recommendations: a requirement that the investigation of misconduct complaints and their adjudication be carried out by separate bodies; a mechanism to review the department's enforcement of misconduct laws; a requirement that the results of misconduct cases be widely disclosed, including the names of accused scientists who have been exonerated; and a requirement that intramural NIH programmes deliver the same assurances and annual reports on misconduct as extramural institutions now have to.

It accepted, however, another controversial proposal requiring institutions to run an educational programme on responsible research conduct and to ensure that all grant recipients attend it.

Meredith Wadman

Sparks fly over climate report

London. Controversy about the rewriting of a key chapter in a United Nations climate change report erupted into open hostility last week when a former president of the US National Academy of Sciences suggested that the rewrite amounted to a "disturbing corruption of the peer review process".

Writing in the *Wall Street Journal*, Frederick K. Seitz, a physicist who is a past chairman of the advisory board of the Strategic Defense Initiative and now president of the George C. Marshall Institute, a conservative science policy organization, attacked the Intergovernmental Panel on Climate Change (IPCC).

Seitz repeated claims made last week by the Global Climate Coalition (GCC), an industry lobby group, in a letter circulated to congressmen and the White House, that the authors had no right to modify the text, which had been agreed after two rounds of government and expert peer review (see *Nature* **381**, 546; 1996).

But he went further by suggesting that the authors' actions, "whatever the intent", effectively deceived both policy-makers and public into believing that science blamed human activities for global warming. Seitz added that the best course of action would now be to "abandon the entire IPCC process".

The IPCC immediately fired off a reply to the newspaper. Its officials say that Seitz did not once contact the IPCC to get "both sides of the story", but simply rehashed arguments put forward by the

GCC. Meanwhile, 40 authors and contributors to the IPCC report have written separately to the *Wall Street Journal* defending the IPCC and Ben Santer, an atmospheric scientist at the Lawrence Livermore National Laboratory in California who acted as the 'lead author' of chapter eight, from charges of acting improperly.

Santer and his co-authors continue to argue that the changes improve the "scientific clarity" of the chapter and that the conclusion — suggesting evidence for a "discernible" human influence on global climate — remains unchanged from both the draft and the final, published text.

But the GCC remains unconvinced. John Shlaes, executive director of the GCC, says the authors altered the science of the report and removed an entire summary, resulting in undue emphasis on a human role in global warming.

Santer says the IPCC rules of procedure state that full reports — as opposed to summaries — do not need to be "approved in detail", implying that late modifications by lead authors are allowed. He argues that this view was endorsed by the State Department, which, in a letter to the IPCC dated 15 November 1995, said: "In keeping with past practice in working group I, it is essential that the chapters not be finalized prior to completion of discussions at the IPCC working group I plenary in Madrid, and that chapter authors... modify the text in an appropriate manner following discussion in Madrid."

Ehsan Masood

New human genetics committee will advise British government

London. The British government has agreed to set up an advisory commission on human genetics in response to a report from the House of Commons Select Committee on Science and Technology, even though it initially rejected the committee's proposal for a statutory committee (see *Nature* 381, 10; 1996).

The new commission will review all aspects of genetics research, as well as its implications for health, insurance, employment and patents. It will also advise on ways of building public confidence in and understanding of genetics, and keep an open door to views from the public. The commission will report chiefly to ministers in the Departments of Health and of Trade and Industry.

The commission will include senior figures from genetics research, industry, public health and possibly the media. It will be expected to take a "broad perspective on the implications of genetics, as distinct from the interests of particular groups", according to a government statement issued on Monday. □

Plant gene resources reviewed

Rome. Delegates from more than 160 countries and 100 non-governmental organizations met in Leipzig, Germany, earlier this week for the twentieth International Technical Conference on Plant Genetic Resources. The week-long conference has been convened to consider and ratify the Global Plan of Action for the Conservation and Use of Plant Genetic Resources for Food and Agriculture.

The United Nations Food and Agriculture Organisation is recommending a 20-point action plan estimated to cost between US\$1.3 billion and \$3 billion over ten years. The plan aims to ensure the "conservation of plant genetic resources as a basis for food security,

to promote better utilization of plant genetic resources, and to promote a better sharing of the benefits of plant genetic resources with countries, communities and farmers". The plan will also make the world's 1,308 gene-banks, which collectively store more than six million seed varieties, more secure and effective. □

Inquiry into Ariane failure

Paris. The European Space Agency (ESA) and the French space agency CNES have set up an inquiry into the failure of Ariane-5 flight 501, which was destroyed less than one minute into its maiden flight two weeks ago, after the launcher veered off course. The nine-member board of inquiry is expected to deliver its report by the middle of next month. In addition, ESA and CNES have decided to strengthen the investigatory powers of the Launcher Qualification Review for all launcher elements. □

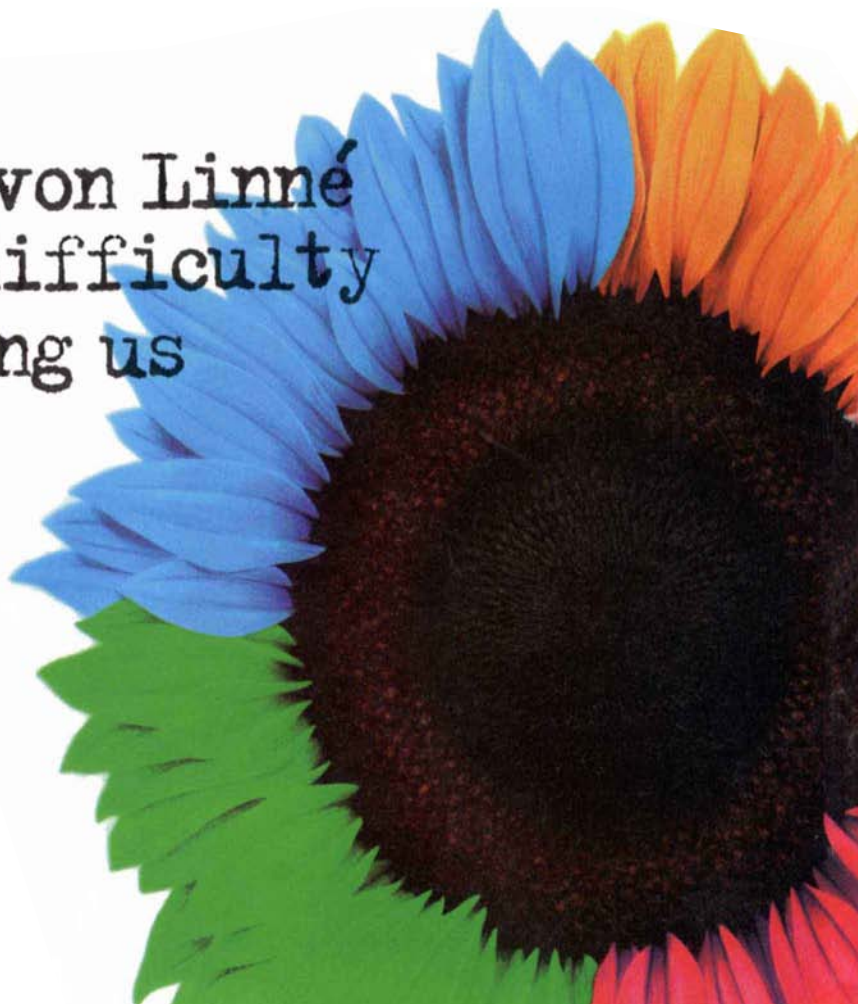
Hughes seeks researchers

Washington. The Howard Hughes Medical Institute (HHMI), the largest philanthropic supporter of research in the United States with assets of \$9 billion and outlays last year of \$410 million, has opened a competition to appoint 25–40 new investigators. Nominations for the elite biomedical research positions are invited by September this year, with appointments expected by May 1997. HHMI currently has 280 investigators working at 62 different institutions, in cell biology, genetics, immunology, neuroscience and structural biology. □

Indian industry ties 'weak'

New Delhi. A three-year United Nations study of India's educational institutions and industrial establishments has found the link between the two to be weak enough to cause "a major dysfunctioning of the science and technical education system". To avert this

Even Carl von Linné
would have difficulty
classifying us



Plant patent quagmire

SIR — A small inaccuracy in your story on plant patenting (*Nature* 381, 178; 1996) makes you too pessimistic about prospects for resolving the problem without amending the European Patent Convention.

The Plant Genetics Systems (PGS) decision of the Technical Board (T356/93) caused concern not merely because it was not the result industry needed, but also because of conflict with earlier decisions. PGS was thought to say that claims that 'embraced' plant — or animal — varieties were invalid. But any claim to a plant or animal will normally 'embrace' a variety of that plant or animal. A quite different test was stated in 'Oncomouse' (T19/93): "Is the subject-matter of the application a variety?" In response to this, the resident of the European Patent Office put a question to the Enlarged Board, which has jurisdiction to resolve such conflicts.

But the Enlarged Board refused the question. In a subtle judgement (G3/95), it explained the basis of the Technical Board's decision. Two points had been argued against the claims (it said):

(1) The attacked claims were bad because they 'embraced' plant varieties.

(2) "Claim 21 defines plants (whether or not they are 'plant varieties' in the sense of the UPOV Convention before they are genetically transformed) which have been genetically modified so that they are herbicide-resistant. This characteristic of genetic herbicide resistance is distinctive and stable in succeeding generations of the plants. *Thus the claimed genetic modification itself makes the plants 'plant varieties' in the sense of the revised UPOV Convention, 1991....*" (Emphasis added.) (UPOV is the Union pour le Protection des Obtentions Végétales, the Union for the Protection of Plant Varieties.)

Although the decision was thought to be based on ground (1), in fact (said the Enlarged Board) it was based only on ground (2). Ground (2) is a new point: so there is no conflict with earlier decisions and the Enlarged Board has no standing to comment.

You are wrong, therefore, in saying that the Enlarged Board confirmed the PGS decision. Nor is this a mere quibble — for in fact the decision in G3/95 completely undermines PGS. Ground (2) — on which the Enlarged Board was careful to make no comment — is clearly untenable. The idea that introducing a single stable gene into any plant material whatever must necessarily produce a UPOV plant variety could only be entertained by someone quite unfamiliar with UPOV practice. UPOV varieties cannot be defined by a single gene: essentially they are defined by all their genes (or at least by all genes that express

readily observable characters). If this is pointed out to a Technical Board in some suitable new test case, there should be little further difficulty in obtaining claims to genetically modified plants.

Such a result would make sense. Recent decisions of the European Patent Office on this point fail to consider why plant varieties are excluded from patent protection under the European Patent Convention. This surely was because there was in place a system for protecting plant varieties (UPOV) that was considered more appropriate: and it was felt necessary to shelter this system from competition. If so, there is no reason to deny protection to generic inventions: these cannot possibly be protected under UPOV.

A claim to a plant variety is a claim to a specific subspecies, whereas a claim to a plant containing a transgene is generic. Generic protection is what biotechnologists require — and this requires patents, rather than plant variety rights, useful though the latter are to protect the products of conventional breeding. Biotechnologists who could only obtain protection for particular plant varieties would be in much the same position as authors who could only prevent copying of their books when set in the original typeface.

Tim Roberts

(Chairman, Biotechnology Committee)
Chartered Institute of Patent Agents,
Staple Inn Buildings,
High Holborn,
London WC1V 7PZ, UK

Address for correspondence: 13 Spring Meadow, Bracknell, Berks RG12 2JP, UK; Email: 70374.26@compuserve.com

Chernobyl legacy

SIR — I (and others) detect a faint bias in your leading article "Chernobyl's legacy to science" (*Nature* 380, 653; 1996). You suggest that the true picture is less clear than in fact it is and that the probability is that the consequences will get worse.

Responsible experts and authorities accept that the total number of deaths resulting from the nuclear accident at Chernobyl over the past ten years is 48; three people died immediately after the accident, 28 within days or weeks, 14 in subsequent years (although those included at least two cardiac failures and one car crash) and, more recently, three children have died from thyroid cancer.

There was general agreement about these figures at a meeting of the European Commission and the Ministries of the Republics of Belarus, Russia and Ukraine in Minsk in March, and at a United Nations anniversary conference in Vienna in April.

You quote Ukrainian "health officials" as saying there have been 125,000 deaths, yet neither the prime minister of Ukraine, the Minister for Chernobyl Affairs, Volodymyr Kholosha, nor any of the other Ukrainian scientists and doctors attending the two meetings, dissented from the figure of 48 deaths. Indeed, Kholosha observed that the widely reported figure of 125,000 deaths in the contaminated areas since 1986 was in fact from all causes, and statistically to be expected.

Two of the three republics, Belarus and Ukraine, are known to be using exaggerated figures in the hope of attracting overseas funding to help to resolve their economic problems. The result is to distort public perception of nuclear power generation.

Eric Voice

25 Miller Place,
Thurso,
Caithness KW14 7UH, UK

Water can be fattening

SIR — Daedalus's recent scheme (*Nature* 381, 118; 1996) for refolding pathogenic proteins *in vivo* rests on the notion that heavy water is only toxic by incongruity with the ordinary kind. Not so: because deuterium weighs twice as much as hydrogen, a bond between deuterium and a heavy atom vibrates with $1/\sqrt{2}$ the frequency of a proton-heavy atom bond. Having lower energy to start with, it takes more energy to break. Administered slowly or instantly, heavy water would act not as a "penetrating oil" on Daedalus's "heavy man", but as sand in his gears.

Deuterium's subtle toxicity might instead make for a suspenseful entertainment. Imagine an important figure, perhaps a head of state, lying in hospital under tight security. An assassin might replace the victim's intravenous feeding solution with one identical but made with heavy water. As the victim became increasingly deuterated he would sink into a deathly torpor. The crime would be averted at the last moment only by the sharp eye of the attending physician, wondering why the patient should have gained 5 or 10 kg on a liquid diet.

David Schweisguth

Department of Molecular
Biology and Biophysics,
Yale University,
PO Box 208114,
New Haven, Connecticut 06520-8114, USA
dcs@proton.chem.yale.edu

Morgagni and the impact factor

IN the letter from M. R. Bonati and A. G. Drusini (*Nature* 381, 271; 1996), G. B. Morgagni's year of birth was incorrect: it should have been 1682. □

Cell-cycle control and its watchman

Tyler Jacks and Robert A. Weinberg

The mammalian cell cycle is controlled by systems that monitor the cell's health and act through the tumour-suppressor p53 and its effector, p21. Therapeutic intervention can restore order to tumour cells lacking p53.

THE mammalian cell-cycle control apparatus and its important regulator, p53 protein, have generated a cottage industry that is rapidly growing to industrial proportions. This research is so attractive because it resolves many long-standing mysteries of cancer cell biology, some of them connected with the responses of tumour cells to therapy. Moreover, p53 is one of the most commonly mutated genes in human cancer, being inactivated in almost half of all tumours¹.

On page 713 of this issue², Waldman *et al.* reveal yet another side of p53's multifaceted activities by concentrating on one of its important effectors, p21. Most work on p53 has centred on its control of the G1 phase of the cell cycle, which falls between the end of cell division (M phase) and the start of DNA synthesis (S phase). But this report suggests that the coordination of the S and M phases of the cycle may also fall within its purview (see figure).

All cell-cycle transitions are initiated by a clock apparatus³. The clock receives a wide variety of growth-controlling signals, integrates and processes them, and then issues marching orders to the cell, which responds by advancing through its growth cycle, by differentiating or by entering into a quiescent growth state, G0.

In fact, the information flux is more complex because the clock apparatus has to monitor compliance with its orders, a form of quality control. For example, if the clock is promoting an advance through S phase, it needs to be informed by the DNA-synthesis apparatus that S phase has been completed before it orders entry into G2 and then M. It also requires reassurance that other intracellular conditions are propitious for cell-cycle advance, for example the energy metabolism and biosynthetic machinery must be in working order. Perhaps most important, the clock must be informed whether or not the cell's genome is intact. The feedback controls made possible by these retrograde signals are often termed 'check-points'⁴. Damage to genomic DNA, lack of oxygen (hypoxia), oncogene activation, and infection by certain viruses represent four distinct physiological signals that generate a negative feedback, resulting in a shutdown of the clock's forward motion.

In principle, these four conditions might be monitored through four distinct, parallel sensing and signalling pathways,

each of which feeds back to the clock in its own way. But vertebrate evolution has apparently devised another scheme. Many of the signals that monitor the state of the cell converge on p53. Once informed by these diverse signals, the p53 protein may signal the clock to shut itself down, usually until some cellular malady is corrected⁵. If, however, the news received about the

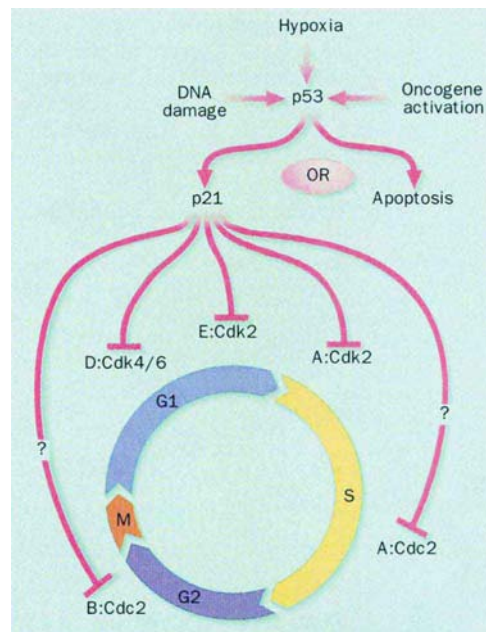
(CDKs) that are active in different phases of the cell cycle (Cdk4, Cdk6, Cdk2) and are central functional components of the clock apparatus^{8,9}.

Unfortunately, this parsimonious design also has a major disadvantage. Like a small corporation that is highly dependent on the talents of a single omniscient executive, the cell is extremely vulnerable to the loss of p53. Inactivation of p53 occurs frequently during the genesis of a wide range of human tumours¹, so the cell-cycle clock becomes oblivious to danger. In p53-mutant cells, the clock continues to issue orders to proliferate when cellular conditions would normally dictate otherwise.

Only recently have we begun to appreciate the consequences of this breakdown in signalling. One apparent price of p53 loss may be an increase in the mutability of the cell's genome¹⁰. In normal cells, a variety of sensors continuously monitor genomic integrity. In the event of damage, perhaps even one double-stranded DNA break per cell¹¹, p53 is alerted and shuts down the clock. The full range of genetic changes (including changes in chromosome number) that can activate p53 is not yet clear. But the rationale of all this is obvious: by shutting down cell-cycle advance (for example in G1), the DNA-repair apparatus is given time to erase genetic damage before S phase is initiated. So a p53-imposed pause in G1 reduces the likelihood of inadvertent replication of mutant DNA.

This explains why p53 has been dubbed the 'guardian of the genome'¹². Indeed, p53-mutant cells are much more prone to accumulate amplified DNA copies^{13,14}, although increases in other types of mutations have yet to be documented. Such increased mutability may accelerate the genetic changes that seem to be the rate-limiting steps in tumour progression.

Mutations in p53 have long been associated with the process of immortalization, in which cells escape apoptosis and go on to create tumours. Recent discoveries indicate that the mortality of normal cell populations, which is manifested in cell senescence and crisis following ex-



The master watchman, p53, receives signals from various monitors of the cell's health. It responds by sending out signals either to trigger apoptosis or otherwise to halt the cell-cycle clock through the cyclin-dependent-kinase inhibitor p21. p21 inhibits the different cyclin-dependent kinases (CDKs 2, 4/6 and Cdc2), working in conjunction with various cyclins (A, B, D, E) to alter the activity of key proteins controlling entry into the different stages of the cell cycle.

cell's health is especially grave, p53, together with other decision-makers, may initiate a more drastic response, causing the cell to kill itself (apoptosis)⁶.

This design of the cell's decision-making circuitry is economical and logical, because it allows a single decision-maker to process a number of distinct incoming messages and, in response to them, emit common growth-inhibitory signals to the cell-cycle clock. Once alerted by various afferent signals, the p53 watchman induces synthesis of a suite of proteins, including p21 (ref. 7). The latter is capable of shutting down cyclin-dependent kinases

tended rounds of cell division *in vitro*, is triggered by collapse of the cell's telomeres¹⁵, the 'caps' on the ends of chromosomes. Mutation of p53 may allow a cell population to ignore the DNA damage associated with telomeric collapse and proceed to a stage where it can resurrect its long-repressed telomerase, the enzyme that keeps the telomere in trim, and thereby achieve immortalization.

Activation of oncogenes, mutant genes that can turn a cell cancerous, also often provokes a p53-mediated growth arrest and apoptosis¹⁶. This may represent an important organismic defence mechanism: cells that have an activated oncogene as a result of some accidental mutation can be consigned to a rapid death by p53. Conversely, p53 mutation may allow an incipient cancer cell to avoid oncogene-induced apoptosis and progress further in tumorigenesis.

The cell's defences against attack by DNA tumour viruses are closely related to those against cellular oncogenes. Oncogenes brought into the cell by viruses (for example, the adenovirus E1A oncogene¹⁷) also provoke a rapid apoptotic response, which should abort the viral replicative cycle by depriving it of a viable host in which to replicate. To ward off this eventuality, the DNA tumour viruses have developed anti-apoptotic functions, including the means to rapidly inactivate p53 (ref. 16). This spares the life of the host cell for long enough to allow the viral lytic cycles to reach completion (and resolves the long-standing puzzle of why DNA tumour viruses, most of which normally engage in rapid, cytolytic growth cycles, happen to encourage cell immortalization of certain host cells in which they cannot replicate).

Most recently, hypoxia has been added to the list of physiological signals that provoke the p53 watchman, causing it to halt cell proliferation and trigger apoptosis¹⁸. This reveals yet another advantage conferred on evolving tumour cells by inactivation of p53. In tumour masses that have a poor blood supply, individual cancer cells can survive under conditions that would lead to the demise of their counterparts with normal p53 function.

The paper by Waldman *et al.* points to yet another fall-out of p53 loss. It focuses specifically on p21, a critical effector of the DNA-damage response orchestrated by p53 (refs 19, 20). These workers demonstrate that tumour cells lacking p21 function undergo multiple rounds of S phase following exposure to various DNA-damaging chemotherapeutic drugs and radiation, without apparently passing through mitosis. So the normally strict control mechanism, which ensures that cells must successfully complete cell division before initiating another round of DNA synthesis, is disrupted. Waldman and colleagues interpret this to indicate

that p21 (and by extension p53) is involved in the normal coupling of the S and M phases of the cell cycle, ensuring that an S phase is not reinitiated before the cell has passed through a preceding M phase. In the p21-mutant cell line tested, the aberrant regulation of the cell cycle was followed ultimately by cell death through apoptosis. Importantly, Waldman *et al.* show that certain p53-mutant tumour lines also exhibit this defect.

An alternative — and equally intriguing — interpretation is that p53 and p21 are responsible for regulating a G2/M checkpoint, blocking entrance into M phase in the face of chromosomal damage or disarray in the mitotic apparatus. This behaviour may be related to observations that cells lacking p53 function fail to halt the cell cycle in response to microtubule-antagonizing antimitotic drugs²¹. The drug treatments used by Waldman *et al.* may disrupt the mitotic process and, in the absence of p53/p21, result in the cell's passing through repeated, highly aberrant mitoses in which chromosomal segregation and cytokinesis, the stage at which a cell's cytoplasm is divided into two, do not occur properly.

Whatever the mechanism behind these

observations, the results of Waldman *et al.* have clear implications for cancer therapy, because tumour cells with p53 mutations may exhibit a selective sensitivity to DNA-damaging drugs. At the same time, the new work suggests the possibility that these same agents may be responsible for the extra or missing chromosomes in those tumour cells in which apoptotic pathways (including p53-dependent ones) have been abrogated.

As the mechanisms of action of p53 continue to be uncovered, the importance of restoring order or inducing apoptotic death in p53-mutant tumour cells becomes ever more apparent. The report by Waldman *et al.* adds to the growing evidence that the state of p53 represents a central determinant of the cancer cell's responsiveness to therapy, and that study of this protein holds great promise for those intent on developing new types of antitumour therapies. □

Tyler Jacks is at the MIT Center for Cancer Research, Department of Biology, MIT, and the Howard Hughes Medical Institute. Robert A. Weinberg is in the Department of Biology, MIT, and the Whitehead Institute, Cambridge, Massachusetts 02139, USA.

- Greenblatt, M. S., Bennett, W. P., Hollstein, M. & Harris, C. C. *Cancer Res.* **54**, 4855–4878 (1994).
- Waldman, T., Lengauer, C., Kinzler, K. W. & Vogelstein, B. *Nature* **381**, 713–716 (1996).
- Reed, S. I. *Trends Cell Biol.* **2**, 77–81 (1992).
- Hartwell, L. & Weinert, T. *Science* **246**, 629–634 (1989).
- Hartwell, L. H. & Kastan, M. B. *Science* **266**, 1821–1828 (1994).
- Lane, D. P. *Nature* **362**, 786–787 (1993).
- el-Deiry, W. S. *et al. Cell* **75**, 817–825 (1993).
- Xiong, Y. *et al. Nature* **366**, 701–704 (1993).
- Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. *Cell* **75**, 805–816 (1993).

- Kastan, M. B. *et al. Cell* **71**, 587–597 (1992).
- DiLeonardo, A., Linke, S. P., Clarkin, K. & Wahl, G. M. *Genes Dev.* **8**, 2540–2551 (1994).
- Lane, D. P. *Nature* **358**, 15–16 (1992).
- Livingstone, L. R. *et al. Cell* **70**, 923–935 (1992).
- Yin, Y. *et al. Cell* **70**, 937–948 (1992).
- Wright, W. E. & Shay, J. W. *Trends Cell Biol.* **5**, 293–297 (1995).
- White, E. *Genes Dev.* **10**, 1–15 (1996).
- Debbas, M. & White, E. *Genes Dev.* **7**, 546–554 (1993).
- Graeber, T. G. *et al. Nature* **379**, 88–91 (1996).
- Deng, C. *et al. Cell* **82**, 675–684 (1995).
- Brugarolas, J. *et al. Nature* **377**, 552–557 (1995).
- Cross, S. M. *et al. Science* **267**, 1353–1360 (1995).

ANTARCTICA

A great lake under the ice

J. Cynan Ellis-Evans and David Wynn-Williams

FOUR kilometres under the ice of central Antarctica is an enormous lake, 200 km long, covering an area of 14,000 km², and in some places at least 500 m deep. These proportions are reported on page 684 of this issue¹ by Kapitsa and colleagues. Lake Vostok, as it is now called, may be a unique habitat for ancient bacterial life. Around 70 other sub-glacial water bodies are known to exist under the central Antarctic ice sheet, so this lake may be part of a vast hydrological system².

An international team has been drilling through the ice sheet near the Russian Vostok station for the past three years, to obtain a glaciological record of climate change. The borehole would in due course break through into the lake, so a workshop was held at the Scott Polar Research Institute in Cambridge to discuss the

glaciological and biological implications of this discovery. The recommendation was that although further physical, chemical and microbiological studies of the lower reaches of the ice core (where the ice is about 500,000 years old) will be highly valuable, it would be imprudent to drill with existing equipment into the lake itself, given the risk of contaminating a uniquely pristine system. It was also recommended that the drilling stop at least 25 m above the lake, which should be enough to prevent drill fluids permeating through. The latest news is that drilling has stopped some 250 m above the lake, and the borehole has been capped, because the Russians have had to close Vostok for the winter.

Lake Vostok lies below a flowing ice sheet, and the melting point temperature at the upstream end, -3.15°C , is sub-

Water kept liquid by warmth from within

It has been known for over 20 years that water can collect in spots beneath the Antarctic ice sheet as it does under smaller, warmer glaciers, and that a subglacial lake exists near Vostok station in central East Antarctica. The occurrence of pools of liquid water beneath the frigid East Antarctic ice sheet, whose mean temperature in its upper parts is -50°C or colder and whose top surface never reaches melting point even on the warmest of summer days, is not as surprising as it might seem.

The reason is simple — the bed of the ice sheet is heated from below by the geothermal flux that characterizes the Earth everywhere, and the four-kilometre-thick blanket of ice, a good thermal insulator, isolates it from the extreme cold above. The temperature of the basal ice is raised to melting point and any excess heat (another source is the frictional heat of sliding) will go towards melting the ice.

As also discussed in the accompanying article by Ellis-Evans and Wynn-Williams, Kapitsa *et al.*¹ have brought together data collected by several techniques over the past four decades to provide a fuller description of Lake Vostok. Kapitsa's re-examination of his seismic soundings made at Vostok station in 1959–60, and new seismic measurements completed just last season, have revealed the great water depth; radio-echo sounding by Robin in the 1970s mapped the ice thickness; and up-to-the-minute analysis of ERS-1 satellite radar altimetry by Ridley provides surface elevations of such remarkable precision that they can prove, from comparisons of surface slope and ice thickness, that the ice floats on fresh water, not salt.

It is the sheer size of Lake Vostok, comparable in area to Lake Ontario and in depth perhaps even to Lake Baikal, that makes this new description so startling. How has so much water collected in one place? Where does it come from? Where does it go?

The answers to these questions are

IMAGE
UNIMAGEABLE
FOR A COPYABLE
FOR A COPYABLE
REASONS

White blue yonder — Vostok station (bottom), almost lost in the immensity of Antarctica.

not yet available in detail, but the general principles of subglacial water movement are easily understood. As in more familiar examples on the Earth's surface, water flow under an ice sheet is simply driven by gravity. The driving force arises from a combination of two factors: pressure ('head') differences, from the loads of different thicknesses of overlying ice, which would force

water to flow even along a horizontal bed (like water from a reservoir moving through a level pipe); and bed-surface slope, down which the water tends to flow just as it does down a mountain-side. Less obvious is the fact that surface slopes (head gradients) are ten times more effective at driving water than bed slopes. Consequently, in most unfrozen places under most glaciers, water flows along the bed in the general direction of the surface slope.

Ice sheets in general, however, and particularly in central East Antarctica, are characterized by minuscule surface slopes — 1 metre per kilometre is typical. It follows that bed slopes of as little as one degree will control the water flow, so the water will collect in depressions in the bed, just as it does on the surface. Also, just as on land, it will fill a depression until it reaches the level of an outlet, through which it will then flow. When the width of a lake becomes many times the overlying ice thickness, the ice floats freely, in hydrostatic equilibrium, on (or in) the water of the lake. It then becomes exceedingly flat and level, like an Antarctic ice shelf, and no longer contributes to the driving force.

Kapitsa *et al.* estimate a mean residence time of water in the lake of some 50,000 years; even more intriguing than that is the age of the oldest water, which depends crucially on the deep circulation within Lake Vostok. What that is no one will know before it is measured directly — and there's the rub, as Ellis-Evans and Wynn-Williams explain.

Charles Bentley

Charles Bentley is at the Polar Research Center, University of Wisconsin, Madison, Wisconsin 53562, USA.

stantially different from that at the downstream end, -2.46°C , because of the 500-m drop in height. This temperature differential may induce circulation within the lake, a fact of considerable biological relevance. The downward movement of ice into the melting zone will release compressed atmospheric gases, such as oxygen, trapped in the ice structure for hundreds of thousands of years. In certain permanently ice-covered lakes of southern Victoria Land, continuous exclusion of gases during ice formation at the under-ice surface, coupled with extremely low microbial consumption of these gases, results in supersaturation of both oxygen and nitrogen³. A similar exclusion process may feasibly occur at the downstream edge of the lake/ice-sheet interface in Lake Vostok, so supersaturation is a slight possibility. High oxygen

tensions are known to be toxic to certain microorganisms.

Microbiological studies of the Vostok ice core have revealed a great diversity of microbes, including yeasts and actinomycetes (with antibiotic-synthesizing potential) which remain viable in ice for up to 3,000 years, and viable mycelial fungi up to 38,600 years old. A unicellular alga, *Crucigenia tetrapodia*, has been found 1,525 m down (in ice about 110,000 years old)⁴ and diatom shells have been observed at 2,375 m (180,000 years old; S. S. Abyzov, personal communication). The oldest viable forms obtained to date are spore-forming bacteria at 2,395 m (200,000 years old)⁵. There is therefore a ready source of microbes seeding the lake.

Radar altimetry and radio-echo studies suggest that the lake water has little or no

salts in it. The possibility does exist, however, of inorganic ions of atmospheric origin, concentrated in the ice, creating an acid lake environment. The likely dearth of nutrients does not necessarily mean that this is a sterile environment, as microbes tend in such circumstances to shrink and form microcells ($<0.1\ \mu\text{m}$ diameter), and can slow down their metabolism⁶. The high pressures to which the microbes will be exposed on melting of the surrounding ice will affect the viability of some microorganisms, but pressure resistance is well documented in marine microbes⁷.

It appears that water temperatures are just below zero, posing few problems for Antarctic microbes, and the geothermal heating that keeps the lake liquid (see the accompanying article by Bentley) could create an even warmer environment in the

sediments, up to 100 m thick, that we believe to be present. Microbes released from the ice sheet will sediment out, and so it is the lake sediment deposits that should provide the most interesting material for microbiologists. The sediments are a potentially valuable chronological record in themselves, and offer solid surfaces and a source of mineral ions for microbes to use.

When justifying microbiological studies of Lake Vostok, comparisons with research in abyssal marine environments are inevitable. But the lake has unique features, including the timescale for material to reach the lake sediments, which is far longer ($\sim 10^5$ years) than that for the deep sea. Also, the biology of Lake Vostok will be entirely microbial, whereas the marine environment has food chains of much greater biodiversity. The prime biological value of Lake Vostok lies in its microbes' gene pool, which should contain characteristics developed perhaps 500,000 years ago, and which has since been isolated from both regional and global changes.

This pool could be readily studied using modern polymerase chain reaction techniques and could perhaps yield substances of medical or industrial application, such as enzymes or antibiotics. Modern-day Antarctic microbes are already being examined in this context. Just as the discovery of the hot 'black smoker' areas in abyssal marine environments has provided new insights into microbial physiology, investigations of microorganisms surviving in an isolated, permanently cold and nutrient-poor environment under high pressure could potentially be very profitable.

Obtaining microbiological material from Lake Vostok would be technically difficult and very expensive, given the location and the absolute need to maintain its pristine nature, but an engineering solution is probably feasible, given time and international commitment. The first priority is to continue the non-intrusive geophysical exploration of the lake begun this austral summer. The biology will have to wait a little longer. □

J. Cynan Ellis-Evans and David Wynn-Williams are at the British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK.

At the magma ocean floor

Michael Walter

GLOBAL warming had an entirely different meaning early in Earth's history, when changes in sea level depended on how much rock was melted into an ocean of magma. But cooler days prevailed, and the planet that emerged more than four billion years ago from this fiery era, aptly named the Hadean, was a differentiated body, with a metallic core and a solid silicate mantle. This differentiation should have left an indelible geochemical signature, but signals from upper mantle rocks have proved curiously difficult to decipher. There is no unequivocal geochemical evidence that there ever was a magma ocean, leading some geologists to believe that it didn't exist. But data presented by Li and Agee on page 686 of this issue¹ show that both the absolute and relative abundances of two key elements, nickel and cobalt, in rocks from the mantle could have been produced by equilibrium between silicate melt and metallic melt at a pressure of about 28 GPa (280,000 atmospheres), corresponding closely to the boundary between the upper and lower mantle.

Elements that partition strongly into metallic phases, such as nickel, cobalt, rhenium and platinum, are called siderophilic, or iron-loving. During differentiation they were almost totally scavenged into Earth's core, but a little of each remained in the silicate mantle. If we can measure the abundances of siderophilic elements in the mantle and have an understanding of their partitioning behaviour over a wide range of pressures, temperatures and compositions, and if metal segregation was an equilibrium process, it should be possible to deduce where and how the core formed. That would tell us much about how the Earth accreted and what the planet was like early in its formation, a period about which we know very little but which set the stage for the subsequent four billion years of the planet's evolution.

Attempts to reconcile siderophilic element abundances in the mantle with experimentally determined partition coefficients have led to a geochemical conundrum². On the basis of coefficients measured at one atmosphere and 1,500 to 1,900 K, siderophilic elements are overabundant in rocks from the upper mantle by as much as several orders of magnitude^{2,3}. If the accretion of the Earth was a gradual and orderly process, then metal segregation at these low pressures and temperatures may be applicable, but stochastic accretion models predict that most of Earth's mass was added by giant impacts of bodies tens to hundreds of kilometres in diameter⁴. Physical models show that the energy buried in the Earth as a

result of such catastrophic events is enough to have substantially or even entirely melted the proto-Earth⁵, so the core may have segregated from the mantle at very high temperatures and pressures. In response, experimentalists have been pushing their apparatus to the limits in order to measure the behaviour of siderophilic elements under extreme conditions⁶⁻⁸.

Nickel and cobalt have critical roles in this drama because they are both refractory and presumably were present in the proto-Earth in the same proportions as in chondrite meteorites. In samples from the mantle, nickel and cobalt are present at about one-tenth of their chondritic abundances, but more importantly they have an approximately chondritic relative abundance ratio. This places a very tight constraint on models of core formation — the processes involved cannot have fractionated these elements.

At pressures up to 12 GPa and temperatures above 3,000 K, experiments have shown that nickel and cobalt become less siderophilic with increasing pressure and temperature, with the pressure effect more marked⁶⁻⁸. Li and Agee have extended the pressure range to 20 GPa by measuring partition coefficients between metallic sulphide melt and silicate melt. They used Allende CI chondrite as a starting material because it has a primitive composition, close to that of the solar photosphere and probably to that of the nebular material that accreted to form the Earth. They find that the pressure effects continue to higher pressures and are not sensitive to the sulphur content of the metal. By extrapolating their data, the authors predict that the absolute and relative abundances of nickel and cobalt in the mantle can be explained by metal-silicate equilibrium at about 28 GPa.

Li and Agee's results are intriguing for two reasons. First, there was no reason to assume there would be a unique pressure at which both the absolute and relative abundances of nickel and cobalt are satisfied, and second, 28 GPa is about the pressure at the upper-mantle/lower-mantle boundary. The authors speculate on several models that may explain this pressure of equilibration. Liquid metal may have 'rained-out' from the then-molten upper mantle, to pool on top of the solid lower mantle and equilibrate with silicate melt there. Eventually, large metal accumulations a few kilometres across became gravitationally unstable and sank to form the core, undergoing little re-equilibration with lower mantle minerals⁹. Or perhaps 28 GPa is instead an average pressure, in which case metal

1. Kapitsa, A. P., Ridley, J. K., Robin, G. de Q., Siebert, M. J. & Zotikov, I. A. *Nature* **381**, 684–686 (1996).

2. Oswald, G. K. A. & Robin, G. de Q. *Nature* **245**, 251–254 (1973).

3. Simmons, G. M., Vestal, J. R. & Wharton, R. A. Jr in *Antarctic Microbiology* (ed. Friedmann, E. I.) 491–542 (Wiley-Liss, New York, 1993).

4. Abyzov, S. S. in *Antarctic Microbiology* (ed. Friedmann, E. I.) 167–218 (Wiley-Liss, New York, 1993).

5. Lipenkov, V. Y., Barkov, N. I., Duval, P. & Pimental, P. *J. Glaciol.* **35**, 392–398 (1989).

6. Morita, R. Y. *Can. J. Microbiol.* **34**, 436–441 (1988).

7. Jannasch, H. W. & Taylor, C. D. A. *Rev. Microbiol.* **38**, 487–514 (1984).

and silicate could have equilibrated in a melt column extending well into the lower mantle.

To add to this, Richter *et al.*¹⁰ have analysed experimental partitioning data for nickel, cobalt, tungsten, molybdenum and phosphorus, including the data of Li and Agee, and have independently come up with a similar result. The mantle abundances of all these elements can be explained if metal-silicate equilibration occurred at about 27 GPa, but at a temperature (2,200 K) below the anhydrous silicate solidus, leading them to speculate that equilibration occurred in a magma ocean that was actually wet.

At present, there are not enough high-pressure partitioning data for all siderophilic elements to test these models more rigorously. Can mantle abundances of sulphur and oxygen be explained? Can metallic melts 'wet' lower-mantle silicates, allowing segregation by percolation, or will these minerals impose a barrier to melt percolation? What was the role of

partitioning among silicates crystallized from the magma ocean? The important thing is that geochemists now have specific and testable magma-ocean models to aim at — so let the arrows fly. □

Michael Walter is at the Institute for the Study of the Earth's Interior, Okayama University, Misasa, Tottori-Ken 682-01, Japan.

1. Li, J. & Agee, C. B. *Nature* **381**, 686–689 (1996).
2. Jones, J. H. & Drake, M. J. *Nature* **322**, 221–228 (1986).
3. Newsom, H. E. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 273–288 (Oxford Univ. Press, New York, 1990).
4. Wetherill, G. W. *Science* **228**, 877–879 (1985).
5. Benz, W. & Cameron, A. G. W. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 61–68 (Oxford Univ. Press, New York, 1990).
6. Walker, D., Norby, L. & Jones, J. H. *Science* **262**, 1858–1860 (1993).
7. Hillgren, V. J., Drake, M. J. & Rubie, D. C. *Science* **264**, 1442–1445 (1994).
8. Thibault, Y. & Walter, M. J. *Geochim. cosmochim. Acta* **59**, 991–1002 (1995).
9. Stevenson, D. J. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 231–250 (Oxford Univ. Press, New York, 1990).
10. Richter, K., Drake, M. J. & Yaxley, G. *Phys. Earth planet. Inter.* (in the press).

AIDS

Hot fusion of HIV

Robin A. Weiss and Paul R. Clapham

OUR understanding of HIV infection has leapt forward by the coming together of two hitherto unrelated facets of AIDS research, announced by Deng *et al.*¹ and Dragic *et al.*² on pages 661 and 667 of this issue, and by Alkhatib *et al.* in next week's *Science*³. The papers show that in order to infect a cell, not only does HIV have to adsorb to the CD4 receptor on the cell surface, a well-established point of entry, it must also enrol the help of a second receptor known as CC-CKR-5. This co-receptor is a receptor for the attractant molecules (known as β -chemokines) RANTES, MIP-1 α and MIP-1 β .

The new findings follow hot on the heels of two recent seminal discoveries: first, that these same chemokines are the major soluble inhibitors of HIV infection secreted by T lymphocytes⁴; and second, that a protein related to the interleukin-8 α -chemokine receptor serves as the receptor for T-cell-line-adapted strains of HIV-1 that induce formation of syncytia⁵ (fused groups of cells).

The α - and β -chemokines are small chemotactic cytokines that attract different types of leukocyte to sites of inflammation⁶: α -chemokines such as interleukin-8 activate neutrophils, whereas the β -chemokines activate T lymphocytes, basophils, eosinophils and macrophages. The cell-surface receptors for chemokines belong to the large family of seven-transmembrane-spanning (7tm), G-protein-coupled receptors⁷. The new results show that certain receptors of this type

cooperate with CD4 in allowing entry of HIV strains with distinct cellular tropisms, or a preference for the type of cell they infect. But here is a word of warning over nomenclature: some of the HIV strains called macrophage-tropic^{1,3} will equally well infect mitogen-stimulated peripheral-blood T lymphocytes.



Electron micrograph of HIV replicating in T lymphocytes. The arrow indicates the envelope glycoprotein complexes that interact with the cellular CD4 and 7tm receptors. Scale bar, 100 nm. (Figure courtesy of H. Gelderblom.)

The new results resolve two long-standing conundra on HIV infection. The first dates from 1988, when Jay Levy and colleagues reported that CD8-positive lymphocytes release a soluble factor that inhibits HIV infection in a manner not restricted by the major histocompatibility complex⁸. The identification of these factors as RANTES, MIP-1 α and MIP-1 β (ref. 4) led to the observation that HIV-

infected people who do not progress to disease have higher levels of β -chemokines in their blood. Perhaps the strong relationship, just published⁹, between age at HIV infection and progression to AIDS might also be explained by falling β -chemokine titres. Moreover, the discovery led to the demonstration¹⁰ that people repeatedly exposed to HIV infection through 'unsafe' sexual activity with HIV-positive partners have unusually high levels of β -chemokines.

The present studies^{1–3} show that RANTES, MIP-1 α and MIP-1 β all inhibit HIV-1 infection by blocking viral fusion and entry, and that expression of the CC-CKR-5 chemokine-receptor gene together with that for CD4 renders cells susceptible to infection by primary non-syncytium-inducing (NSI) strains of HIV-1 that naturally infect peripheral-blood CD4-positive lymphocytes and macrophages. Furthermore, Koup and Moore's group² shows that individual lymphocyte clones derived from individuals who appear to be resistant to infection produce high levels of inhibitory β -chemokines. It is not yet clear whether the chemokines competitively block their receptors against HIV entry, or whether chemokine binding results in downregulation of cell-surface expression of the receptor.

The second puzzle dates from 1986, when we showed by gene transfer that human CD4 is sufficient for binding HIV-1 to cells but is not sufficient for HIV-1 penetration or for fusion of the viral envelope with the host-cell membrane¹¹. Many groups since then have searched for a second participant or co-receptor for HIV entry. Various candidate molecules were put forward, sometimes with more fanfare than fact, but have not stood the test of critical examination.

Last month, Ed Berger and colleagues⁵ convincingly showed that a 7tm receptor allowed entry and fusion of T-cell-line-adapted, syncytium-inducing HIV envelopes, but was not permissive for primary NSI strains. This receptor, named LESTR, HUMSTR or fusin, is an orphan receptor with no known ligand^{12,13}. Thus, like the multiple membrane-span-

ning receptors for C-type retroviruses¹⁴, its function of permitting viral entry has preceded the identification of the natural ligand.

As the HIV-inhibitory β -chemokines also recognize receptors of the 7tm type, several laboratories raced to see whether primary, NSI HIV utilizes a β -chemokine receptor. The various β -chemokines recognize multiple CC-CKR receptors, but

the finger was pointing to CC-CKR-5 because it specifically binds RANTES, MIP-1 α and MIP-1 β . With the recent publication of its sequence¹⁵, expression cloning was possible and it soon became apparent that CC-CKR-5 serves as the co-receptor to CD4 for fusion and entry of those NSI strains of HIV that infect peripheral-blood lymphocytes and macrophages¹⁻³.

Although two long-standing questions about HIV infection *in vivo* and *in vitro* have now been fused to a satisfactory conclusion, these novel findings are by no means the end of the story. Rather, they are opening up a new area of AIDS research. Other co-receptors are likely to come to light: for example, HIV-2 and simian immunodeficiency virus have distinct cell tropisms from HIV-1 (refs 16, 17). The actual mechanism by which the HIV envelope (see figure) and its cellular receptors trigger membrane fusion¹⁸ re-

mains to be elucidated, and we may look forward to a new class of inhibitors that block this interaction.

The effects of HIV envelope glycoproteins on the signalling by 7tm receptors for chemotaxis, mitogenesis, apoptosis, and the secretion of further cytokines may prove to be important in unravelling HIV pathogenesis in the immune system and the brain. As Feng *et al.*⁵ point out, there is now a possibility of developing transgenic animals for HIV infection that express

both human CD4 and the relevant co-receptor. Animal models may not be a simple solution, however; most mouse cells, for instance, exert post-penetration restrictions on HIV replication¹⁷. Nonetheless, these recent discoveries¹⁻⁵ have set AIDS researchers alight with enthusiasm. □

Robin A. Weiss and Paul R. Clapham are at Chester Beatty Laboratories, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK.

- Deng, H. K. *et al.* *Nature* **381**, 661-666 (1996).
- Dragic, T. *et al.* *Nature* **381**, 667-673 (1996).
- Alkhatib, G. *et al.* *Science* (in the press).
- Cocchi, F. *et al.* *Science* **270**, 1811-1815 (1995).
- Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. *Science* **272**, 872-877 (1996).
- Murphy, P. M. A. *Rev. Immunol.* **12**, 593-633 (1994).
- Neote, K., DiGregorio, D., Mak, J. Y., Horuk, R. & Schall, T. J. *Cell* **72**, 415-425 (1993).
- Walker, C. M., Moody, D. J., Stites, D. P. & Levy, J. A. *Science* **234**, 1563-1566 (1988).
- Darby, S. C., Ewart, D. W., Giangrande, P. L. F., Spooner, R. J. D. & Rizza, C. R. *Lancet* **347**, 1573-1579 (1996).
- Paxton, W. A. *et al.* *Nature Med.* **2**, 412-417 (1996).
- Maddon, P. J. *et al.* *Cell* **47**, 333-348 (1986).
- Raport, C. J. *et al.* *J. Leukocyte Biol.* **59**, 18-23 (1996).
- Wells, T. N. C. *et al.* *J. Leukocyte Biol.* **59**, 53-60 (1996).
- Weiss, R. A. & Taylor, C. *Cell* **82**, 531-533 (1995).
- Samson, M., Labbe, O., Mollereau, C., Vassart, G. & Parmentier, M. *Biochemistry* **35**, 3362-3367 (1996).
- Clapham, P. R., Blanc, D. & Weiss, R. A. *Virology* **181**, 703-715 (1991).
- McKnight, A., Clapham, P. R. & Weiss, R. A. *Virology* **201**, 8-18 (1994).
- Moore, J. P., Jameson, B. A., Weiss, R. A. & Sattentau, Q. J. in *Viral Fusion Mechanisms* (ed. Bentz, J.) 233-289 (CRC, Boca Raton, 1993).

MATERIALS CHEMISTRY

Scandal of crystal design...

Philip Ball

It's how you say it that counts. When in 1988 John Maddox declared in this journal that the inability to predict *a priori* the structure of any crystalline material was a "scandal", he could not have more effectively thrown down the gauntlet. Eight years on, the smarting is still felt — and not least because in large part the scandal remains.

This lack of an infallible predictive tool is perhaps why it is so hard to identify any crystalline material in wide technological use that has been prepared by design. Piezoelectrics such as lead zirconate-titanate (PZT), nonlinear optical crystals such as lithium niobate, the copper-oxide superconductors — all are products of more or less chance discovery. Even zeolites, perhaps the favourite of crystal chemists' 'designer' materials, are either inorganic natural products or artificial variants thereof whose synthetic predictability is imperfectly developed.

What, then, are the prospects for using rational design to make new, useful crystalline materials? If we cannot yet with certainty predict the phase diagram of silica, what hope can there be for a drawing-board approach to engineering the solid state? This was the question faced by participants at a meeting last month*, who showed that rejoinders to Maddox's challenge now make an impressive list. While a set of universal rules for planning crystal structures remains a dream, there is now both a tool kit and something approaching

an instruction manual for building solid-state materials to order.

At present, the price that must be paid is to declare large swathes of the periodic table out of bounds. Metals have some very appealing uses in crystal design, but

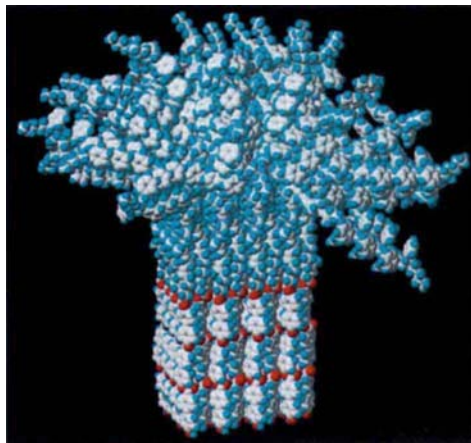


FIG. 1 Supramolecular mushrooms — these aggregates are formed from linear molecules in which a rod-like unit, which packs into an ordered stem, is connected to a flexible unit, which forms the disordered head.

by and large inorganic chemistry as such is still too daunting: prediction there demands a full quantum-mechanical treatment, with little guarantee of success and still less of qualitative guidelines for would-be crystal engineers. That notwithstanding, the ornate frameworks that can be constructed from inorganic building blocks alone are astonishing, and to the classic example of zeolites (whose micro-

porous channels and cavities are made up of tetrahedral silicate and aluminate units) we can now add a dazzling array of microporous vanadate frameworks, where the square pyramidal VO₅ building block is predominant (Bob Haushalter, NEC Princeton).

But the strongest message of the meeting was that, to gain a good idea in advance of what to expect from a crystal structure, you are best off working with organic building blocks. Preformed organic units — called tectons by Jim Wuest (Univ. Montréal) — can make crystal engineering (a term now 30 years old) more a matter of construction-kit assembly than of *ab initio* number crunching. The idea is simple: you design your tectons so that there are only a very few ways, perhaps even a unique way, in which they can join together. If the design is good, and if the tectons are given 'sticky ends', they should self-assemble from (generally) solution, precipitating as the crystal structure of your choice.

Of this strategy, the last step (precipitating good single crystals) remains a black art. But the rest is, in principle, a purely rational matter. A favourite target is the diamondoid lattice: a three-dimensional network of tetrahedral vertices, like the carbon framework of diamond. This should result from tectons designed to self-assemble in tetrahedral geometry — four-armed molecules with each sticky end pointing to the corner of a tetrahedron. Wuest has made tectons of this sort in which the sticky ends form two hydrogen bonds with an identical partner. The diamondoid lattices constructed from these units contain large channels, up to a nanometre-and-a-half wide.

Hydrogen bonds are an attractive

*NATO Advanced Research Workshop on *Self-assembly in Synthetic Chemistry*, Val Morin, Québec, 16-21 May 1996.

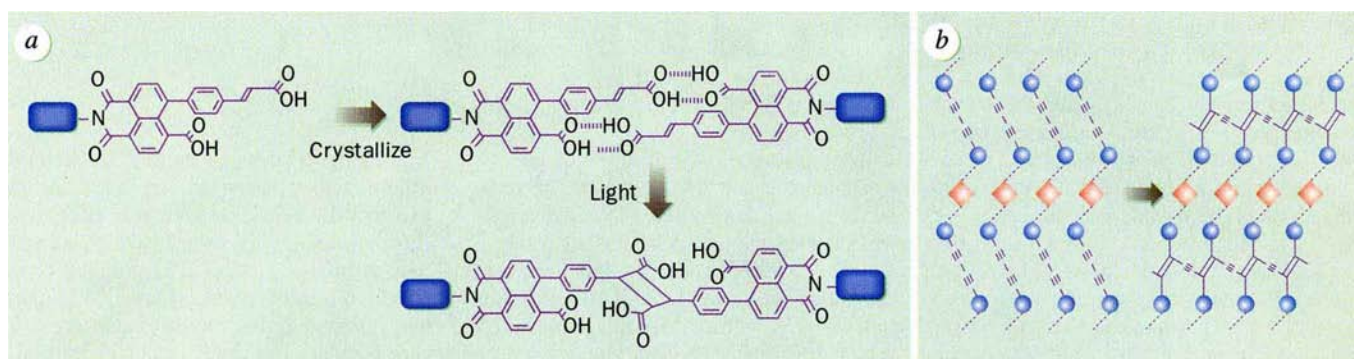


FIG. 2 *a*, Covalent links between self-assembling molecules can be forged by including unsaturated bonds in the 'sticky ends'. Here the appendages of some central molecular hub (blue rectangle) associate via hydrogen bonds in the crystal, and are then crosslinked by irradiation to induce cycloaddition of the adjacent carbon-carbon double bonds.

b, Diacetylenes can be 'pre-organized' for covalent crosslinking by attaching to each end of the linear molecules chemical groups (blue balls) that form non-covalent links (dashed lines) with 'organizing centres' (red squares) in a co-crystal. Irradiation-induced crosslinking then generates one-dimensional polydiacetylene chains.

glue for this building-block approach to molecular crystals, for several reasons. They are directional in space, so that the disposition and orientation of sticky ends directs the topology of the network; they are weak enough (one should perhaps say reversible or 'labile' enough) to allow for annealing and editing during the assembly process, yet strong enough to generate stable crystals; and they are relatively deformable, allowing for some bending strain to be accommodated between the connected tectons (recalling the still greater flexibility of the Si-O-Si linkage, which enables zeolites to adopt such a variety of frameworks). At present, it is fair to say that hydrogen bonds are the most common glue in designed molecular assemblies of all sorts, be they crystals or solution-phase supramolecular complexes.

But greater geometric scope, more variety in bond strengths, and possibly greater potential to design functionality (optical, magnetic, electronic, catalytic) come from the use of coordination bonds to metal ions as the connections between tectons. Richard Robson (Univ. Melbourne) described a bewildering range of open-framework structures made from organic components linked via metal ions at their vertices. Square porphyrin-based tectons joined at the corners via tetrahedrally coordinated copper(I) ions, for instance, provide a tetragonal framework structure with rectangular channels 15 by 20 Å across (see *Nature* **369**, 727; 1994).

Any forces weaker than full covalent bonds might in principle be exploited to direct crystal self-assembly. Even so weak and non-directional an interaction as the van der Waals force has its value when combined with appropriate tecton geometry, as illustrated in the one-dimensional stacks reported by Wais Housseni (Univ. Louis Pasteur, Strasbourg). Here, back-to-back cup-shaped calixarenes ('exoreceptors' by virtue of their divergent binding-site orientations) are held together by linear diyne molecules inserted

into opposed cups. And Sam Stupp (Univ. Illinois) showed that geometric packing effects alone, rather than short-range molecular interactions, can apparently force a molecular crystal into a non-intuitive structure that arguably contains an element of design. Stupp has made 'supramolecular mushrooms', finite-sized aggregates of surprisingly regular size and shape comprised of linear oligomeric molecules with a rod-like section, that can stack in a crystalline manner, and a flexible section that forms disordered aggregates. The molecules aggregate into a close-packed stem of rods and a mushroom head of disordered strands, whose size is ultimately limited (to 100 or so molecules) by the point at which the bulging of the head prevents further packing in the stem (Fig. 1).

These mushrooms form lamellar stacks, visible by transmission electron microscopy, in which one might expect a bilayer stem-to-stem packing configuration. But instead, efficient packing seems to favour a stem-to-head arrangement, giving the lamellar films different surface energies on each face, which is evident from their markedly different contact angles against water droplets.

The problems of a tectonic approach remain substantial, however. Predictability is often evident but far from guaranteed. In general it is made easier by restricting the dimensionality: two-dimensional sheets made from flat tectons are easier to design than three-dimensional networks, and because there is only one way that identical sheets can stack, the third dimension then takes care of itself. But polymorphism is a constant bugbear: for many tectons the intended crystal structure is not unique, and it is hard to know which of several possible connected networks will be preferred.

For crystals that incorporate large voids, meanwhile, abhorrence of a vacuum (more properly, the enthalpic benefit of high packing density) commonly leads to interpenetration: rather than a single network, one obtains several identical,

topologically independent copies which interpenetrate to fill space. This happens for both Wuest's diamondoid nets and Robson's tetragonal nets, for example. If the aim is to achieve large voids or channels (so as to mimic the molecular sieving propensity of zeolites), this is a nuisance. One way of avoiding the problem may be to use tectons whose covalent structure predetermines the channel spaces—rigid, channel-forming rings, for instance, cannot interpenetrate (see *Nature* **371**, 591; 1994).

Another drawback of organic molecular crystals is their inherent lack of stability (thermal, and sometimes mechanical and/or chemical). Haushalter reminded the meeting that, for all the talk one hears of organic zeolites, no one is going to make an open-network organic crystal that can be boiled for hours in crude oil. Indeed, few of these channel-bearing crystals survive removal of the solvent that fills the pores. But improved stability might be obtained by incorporating into the tectons' sticky arms groups that can be covalently crosslinked once weaker interactions have directed the self-assembly. Ken Feldman (Pennsylvania State Univ.) described steps in this direction, whereby hydrogen-bonded dimers could be linked by cycloaddition between two unsaturated bonds in the linkers (Fig. 2*a*). Joe Lauher and Bill Fowler (SUNY at Stony Brook) have used the tectonic approach to pre-arrange diacetylene molecules into two-dimensional sheets in such a way as to facilitate crosslinking (by γ -irradiation) to give polydiacetylene chains (Fig. 2*b*).

Although there is still ample scope for refining the control of structure, the field of crystal engineering already faces a growing need to define some practical targets—in other words, to engineer function too. Here it seems to me that the agenda is set not so much by what might be possible as by what the competition is already capable of. For instance, micro-electronic engineers have for years practised crystal engineering of a rather different kind, in which structural sophis-

tication is secondary to the more important business of engineering electronic properties, namely band gaps. This can be achieved merely by adjusting the lattice spacing of III–V semiconductor alloys through control of composition or by careful introduction of strain into thin-layer, non-lattice-matched superlattices. Molecular-beam epitaxy has brought this technology to the status of a fine art, capable for instance of generating continuous compositional gradients with atomic-layer position. It is doubtful that molecular crystal engineering will easily make much impact here.

Although many such crystals are currently designed to have large pores, it is not clear (for the reasons stated) that there is much future in striving to make these selective industrial catalysts to rival zeolites. Rather, it is in the gentler arenas that inorganic microporous materials have entered — molecular sieving and sensor technologies, for example — that large-pore organic crystals might compete. Particularly attractive here is the possibility of designing helical pores for chiral separation.

In terms of dense materials one might wonder whether self-assembly could be used to pre-organize components for subsequent covalent linkage into materials that are otherwise hard or impossible to make: cubic boron nitride springs to mind, or the mythical β - C_3N_4 that is predicted to have the hardness of diamond. But that is ambitious, to say the least. Better, perhaps, to focus on functional materials such as non-centrosymmetric crystals for nonlinear optics, polar crystals for ferroelectrics and piezoelectrics, or magnetic materials in which the magnetic interactions of metal centres might be tuned by adjusting their symmetry and separation.

In getting function from form, nature is an old hand. But while those practising self-assembly in general draw much inspiration from biology, there are no perfect, structurally elegant molecular crystals to be found there. The materials of nature are composites, and long-range order is rare. In fact, properties and functions are arguably dictated more by defects than by regularities — in structural proteins, the parts of a peptide chain that break a propensity towards crystallization (β -sheet formation, say) are essential for conveying flexibility and preventing brittleness. Defects may turn out to be as important for obtaining useful engineered crystals as regularity, a point emphasized by Sam Stupp. They can, for instance, relieve strain or act as trapping sites for electronic excitations (as in photorefractive crystals). If ever we obtain absolute structural control over the solid state, we might then need to loosen our grip in order to make that facility useful. □

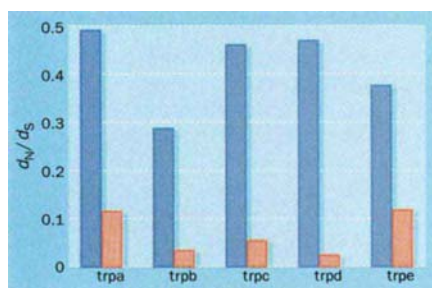
Philip Ball is an associate editor at Nature.

... and scandalous symbionts

Laurence D. Hurst and Gilean T. McVean

THE persistence over millions of years of lineages of organisms that never have sex is, in the words of John Maynard Smith¹, “an evolutionary scandal”. Most theories² do indeed have it that in the long term asexuality will usually lead to extinction, but work by Nancy Moran, published in the *Proceedings of the National Academy of Sciences*³, may help to explain why some asexual lineages persist.

Sex should promote evolutionary change and survival if selective forces fluctuate,



Ratio of mean non-synonymous (d_N) to mean synonymous (d_S) substitutions per nucleotide site for five tryptophan biosynthetic enzymes in endosymbiotic and free-living bacteria. Red, the ratio in the endosymbiotic *Buchnera* from two aphid species. Blue, the ratio in the free-living enterics *Escherichia coli* and *Salmonella typhimurium*.

tuate, as is thought to occur during antagonistic parasite–host coevolution, and it should also prevent the accumulation of deleterious mutations^{4,5}. Models of the parasite–host interaction predict that asexual lineages die out rapidly because they are unable to keep up with faster evolving parasites (one need only look at the consequences of growing monocultures⁶ — take the Irish potato famines of the last century, for example). Genome decay models, on the other hand, emphasize the vulnerability of small asexual populations to the accumulation of slightly deleterious mutations, which make a population smaller and by a positive feedback process can lead to ‘mutational meltdown’⁷.

Moran has studied the endosymbiotic bacteria (*Buchnera*) of aphids. The bacteria are passed from mother to progeny every generation, and have apparently done so for some 100 million years⁸. The absence of biparental inheritance precludes genetic recombination between different clonal lineages, but a bacterial symbiont in an insect host inhabits a cosy world protected from both parasites and environmental fluctuations. So the asexuality of endosymbiotic bacteria has been considered to be consistent with the parasite theory of sex⁹.

These bacteria do, however, pose a problem for the view that sex is a means of

expunging deleterious mutations — populations of endosymbiotic bacteria go through a ‘bottleneck’ when they pass through the eggs of their hosts. This limits the population size and makes them prone to mutational meltdown. Why then have asexual endosymbionts such as *Buchnera* not decayed to extinction? There are two possible explanations. First, the mutational decay they suffer may happen only very slowly; second, they may have accelerated rates of accumulation of deleterious mutations, but accompanied by compensatory responses.

Moran compared the rate of evolution of *Buchnera* genes with that of the same genes in closely related, free-living bacteria that live in huge populations, some of which have recombination. Differences between genes in two lineages may be of one of two types: there are those that have no effect on the amino-acid composition of the protein concerned (synonymous substitutions) and those that do (non-synonymous). Moran could then ask whether the endosymbionts have relatively accelerated rates of non-synonymous substitution, as genome-decay models would predict. She found that this was indeed the case — see figure. The same has been found for one Y-linked gene in mammals^{10,11} (the Y chromosome escapes recombination and also has a small effective ‘population size’, and so is essentially in the same boat as an endosymbiont). Moran also demonstrates that 16S RNA evolves much faster in five independently evolving symbiont lineages than in their free-living relatives.

The high ratio of non-synonymous to synonymous mutations is compatible with genome-decay models. But it could also be due to adaptive change, as could be argued for the Y chromosome. Moran, however, contends that adaptation is not the reason behind the pattern of change seen in endosymbionts, because the percentage of adenine (A) and thymine (T) in the endosymbiont genome is very high and many of the A/T substitutions occur in non-synonymous positions. The high A and T content could reflect adaptive genome-wide selection favouring As and Ts. But selection for base composition would be expected to act primarily on synonymous sites. So most non-synonymous substitutions are probably deleterious.

It seems, then, that the endosymbionts have tolerated accelerated rates of deleterious mutations while avoiding mutational meltdown and extinction. So how have they done it? The rate of accumulation of mutations could be slowed, in the absence of recombination, if the per genome mutation rate could be somehow reduced.

This could be achieved by a reduction of the per locus mutation rate and/or the number of genes in the genome (possibly, as with mitochondria and chloroplasts in eukaryotic cells, losing some to the nuclear genome).

Comparisons of the synonymous substitution rates in free-living and endosymbiotic bacteria so far tentatively suggest that the per locus mutation rate is no different³. And, unlike eukaryotic organelles, *Buchnera* seem to have lost only a few genes⁸. Most notably, however, they have lost the gene (*recF*) encoding a repair enzyme¹². At first sight this seems remarkable, for this would surely seem to be one of the most important genes to hang on to. But again the observation is consistent with theory⁷ — a good way for a clonal lineage to minimize the rate of accumulation of mutations is to ensure that most of them are not slightly deleterious but very harmful, so removing them immediately rather than allowing them to accumulate. Such a strategy is only viable, however, if dead bacteria can be replaced by non-mutated clone members and if bacteria with heavily deleterious genes can never take over a population (for example by fast replication). So symbionts are particularly likely to develop this strategy, being restricted to host cells and surrounded by clonal relatives.

So that is one compensatory trick that *Buchnera* has come up with to avoid mutational meltdown. Another, as Moran points out, is that a few endosymbionts, *Buchnera* included¹³, contain very high concentrations of the chaperone protein GroEL. In free-living forms, chaperones are produced at high temperatures to stabilize proteins that might otherwise denature. Perhaps, Moran conjectures, the endosymbionts employ GroEL during normal existence to stabilize mutationally decayed proteins.

If this is so, the question arises as to whether compensatory mechanisms are peculiar to endosymbionts or are a general characteristic of asexual lineages. If they are widely found, then mutational meltdown need never be a problem. But there are several arguments as to why

these mechanisms are likely to be endosymbiont-specific. These bacteria live in a location where energetically expensive compensation is possible (that is, the host can afford to give the symbionts excess nutrients so that they can produce large quantities of chaperone); where mutations can be easily eliminated by host inter-cell selection; where the host may be able to compensate for some of the symbionts' failings; and where kin selection between members of a clone could favour inter-symbiont synergism.

All in all, then, the fact that long-lasting lineages of endosymbionts are asexual and exist in small populations is not evidence uniquely favouring the parasite hypothesis, and is consistent with both of the main

hypotheses for the maintenance of sex found in most taxa. Moreover, Moran's finding of accelerated evolution lends considerable weight to the importance of mutational decay. But it is not known whether *Buchnera* has reached an equilibrium with respect to the number of deleterious mutations or whether decay is continuing. Perhaps when GroEL itself decays beyond a certain point, the whole system will collapse — for all we know, endosymbionts (and Y chromosomes) may be doomed to extinction. □

Laurence D. Hurst and Gilean T. McVean are in the Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.

CONDENSED-MATTER PHYSICS

Heavy oxygen tips the balance

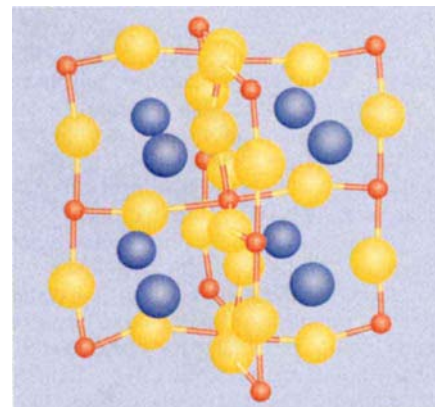
Warren Pickett

IN most materials, magnetic phenomena at room temperature and below are essentially independent of the masses of the nuclei of the constituent atoms. The atoms can usually be considered as infinitely heavy and static in theoretical descriptions of such phenomena, although there are exceptions. Over a decade ago, Kim¹ showed how atomic motion could result in small but measurable changes in the properties of 'itinerant' ferromagnets. It is therefore not so surprising that atomic motion might have a measurable effect on the temperature below which lanthanum-based manganites become ferromagnetic (the Curie temperature, T_c). But the magnitude of the effect, as reported by Zhao *et al.* on page 676 of this issue², is alarming.

Lanthanum-based manganites, such as $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3+y}$, have recently become the subject of much attention by virtue of their unusual — and potentially useful — magnetic properties. At high temperatures they are insulating; at low temperatures they are metallic. But the temperature at which the insulator-to-metal transition occurs can be increased by applying a (large) magnetic field. As a result, if the temperature is held in the region of the transition and a magnetic field is applied, the electrical resistance of the material can be decreased by a factor of 1,000 or more. This phenomenon is now known as 'colossal magnetoresistance' (CMR), and researchers are energetically pursuing its cause. The subsequent observations of structural phase transitions³ and charge order-disorder transitions⁴ driven by an applied magnetic field have only added more questions. The observation by Zhao *et al.*, that simply increasing the mass of the oxygen atoms (by replacing ^{16}O with ^{18}O) decreases T_c by >20 K, now brings some focus to the fundamental

description of these perplexing materials. In particular, it is the first experiment to establish what many have suspected — that atomic motion must be included in any viable description of the manganites.

An interesting complication in the manganites is that T_c coincides with the electronic transition temperature; as the electronic properties switch from insulating to metallic, the manganites transform from being magnetically disordered to magnetically ordered (ferromagnetic).



Structure of the unit cell of LaMnO_3 : the La (blue), Mn (red) and O (yellow) spheres are proportional to their ionic radii. Note that the distortion from a simple cubic perovskite structure is substantial.

This coincidence indicates that the magnetic and electronic properties are strongly coupled. An applied magnetic field, H , will couple mainly to the spin magnetic moment of the electrons (coupling to orbital motion is usually a smaller effect). This coupling is $2S\mu_B H$ (where μ_B is the Bohr magneton). For an $S=3/2$ spin moment in a magnetic field of 5 tesla, the coupling energy is 1 meV or, in tempera-

1. Maynard Smith, J. *Nature* **324**, 300–301 (1986).
2. Kondrashov, A. S. *J. Hered.* **84**, 372–387 (1993).
3. Moran, N. A. *Proc. natn. Acad. Sci. U.S.A.* **93**, 2873–2878 (1996).
4. Hurst, L. D. & Peck, J. R. *Trends Ecol. Evol.* **11**, 46–52 (1996).
5. Crow, J. F. *Dev. Genet.* **15**, 205–213 (1994).
6. Hamilton, W. D., Axelrod, R. & Tanese, R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3566–3573 (1990).
7. Gabriel, W., Lynch, M. & Burger, R. *Evolution* **47**, 1744–1757 (1993).
8. Moran, N. & Baumann, P. *Trends Ecol. Evol.* **9**, 15–20 (1994).
9. Law, R. & Lewis, D. H. *Biol. J. Linn. Soc.* **20**, 249–276 (1983).
10. Tucker, P. K. & Lundrigan, B. L. *Nature* **364**, 715–717 (1993).
11. Whitfield, L. S., Lovell-Badge, R. & Goodfellow, P. N. *Nature* **364**, 713–715 (1993).
12. Lai, C.-Y. & Baumann, P. *Gene* **113**, 175–181 (1992).
13. Baumann, P., Baumann, L. & Clark, M. A. *Curr. Microbiol.* **32**, 279–285 (1996).

ture equivalent, 10 K. Although this is a relatively small energy scale, it makes its presence felt through the strong magneto-electronic coupling — a flip of a spin in an applied magnetic field can change the local electronic structure from insulating to metallic, so affecting the energy balance on a (larger) electronic scale.

Another potentially important contributor to the properties of the manganites is a Jahn–Teller structural instability; the lattice surrounding the Mn^{3+} ions distorts to lift a degeneracy in the energy levels of the electronic ground state of these ions, and so lowers the energy of the entire (electronic+lattice) system. The electronic configuration of the Mn^{3+} ions is d^4 with all spins aligned, with all three t_{2g} orbitals fully occupied, but only single occupation of the two degenerate e_g orbitals (which point directly at the neighbouring oxygen ions).

The ideal structure of these compounds is simple cubic perovskite. But the parent compound of the CMR materials, LaMnO_3 , has a complex structure (see figure), which can be described as a series of distortions from the simple cubic structure: rotation of the MnO_6 octahedra around the z axis; tilting of the octahedra about an axis in the x - y plane; distortion of MnO_6 octahedra resulting in three unequal Mn–O bond lengths; and relaxation of the entire structure to give three unequal lattice constants. The rotational motions result from a misfit of ionic sizes whereas the distortion of the MnO_6 octahedron is widely believed to be the manifestation of the Jahn–Teller instability. As Ca is substituted for La, the distortion of the MnO_6 octahedra nearly vanishes according to diffraction experiments that measure the average structure. But there has been speculation that the Jahn–Teller distortion nevertheless remains, either as an incoherent local distortion (fluctuations in space) or, more exotically, as dynamical Jahn–Teller fluctuations.

Although invoking the Jahn–Teller mechanism to account for data has been common, all phenomena up to this point could in principle be addressed in terms of static atoms. The new result of Zhao *et al.* clearly points to the strong involvement of the dynamic (thermal) motion of the oxygen ions: an increase of the oxygen mass by about 10% drives down T_c by roughly 10%. One of the prime proponents of the importance of Jahn–Teller effects in manganites has been theorist Andrew Millis. His classical model⁵ concentrates on a structural ('Jahn–Teller') transition, rather than on electronic transport and magnetic ordering; as such it does not incorporate the oxygen mass explicitly and therefore does not address the current data.

A simple, and perhaps simplistic, guess at the isotope shift of T_c might go as follows: neglecting lattice fluctuations and their coupling to the electronic and magnetic properties, the transition tempera-

ture is T_{c0} ; accounting for the fluctuations will depress T_c below T_{c0} ; a heavier mass allows smaller fluctuations and will result in less depression of T_c . Thus, T_c increases with mass. Zhao *et al.*, however, report that the heavier mass (M) decreases T_c , so that the isotope shift $\alpha = d\log(T_c)/d\log M$ is negative.

The negative shift, opposing the simple effect of fluctuations, lends credence to Zhao and colleagues' suggestion that the primary effect is of polaronic origin. The Jahn–Teller polaron is a charge carrier accompanied by a local Jahn–Teller distortion; that is, the charge and local distortion effectively move through the material as a single entity (a polaron). It has a kinetic energy W_{eff} that is oxygen-mass dependent, and decreases with increasing mass. Because T_c scales with W_{eff} , the isotope shift is negative. Zhao *et al.* also point out that the variation of the polaronic bandwidth with La/Ca ionic radius can account for the trend to larger isotope shifts for smaller mean ion size on the La–Ca sublattice.

Does this news point to the mechanism of CMR? Not directly, it seems, although it shows that dynamical behaviour is an essential component of the behaviour near T_c . An important question, still open, is this: how can a system with such strong polaronic character near T_c , in which carrier motion is impeded by its strong coupling to atomic motions, become such a good conductor at low temperature? Values of resistivity as low as $10 \mu\Omega \text{ cm}$ have been reported, and this amount of resistivity can be accounted for simply by charge scattering from the randomly distributed La^{3+} and Ca^{2+} ions. The excellent conduction is consistent with simple band metal behaviour, but there is no theory to account for how the strong electron–lattice coupling would 'evaporate' at low temperature.

These new isotope shift results allow one to put a new slant on our picture of the CMR materials. Near T_c , there is a huge increase in resistivity resulting from the heavier oxygen isotope — that is, these are also 'colossal isotoporesistance' materials. It is intriguing to note that the effects of the isotopic mass increase (a lowering of T_c) can, at least approximately, be offset by the application of a magnetic field of the proper strength, which will drive T_c back up to its original value. The essential questions about CMR materials are unanswered, but some of the issues are being clarified. □

Warren Pickett is at the Naval Research Laboratory, 4555 Overlook Avenue, Washington, DC 20375-5000, USA.

1. Kim, D. J. *Phys. Rev. Lett.* **47**, 1213–1216 (1981).
2. Zhao, G.-m., Conder, K., Keller, H. & Müller, K. A. *Nature* **381**, 676–678 (1996).
3. Asamitsu, A. *et al. Nature* **373**, 407–409 (1995).
4. Tomioka, Y. *et al. Phys. Rev.* **B53**, R1689–R1692 (1996).
5. Millis, A. J. *Phys. Rev.* **B53**, 8434–8441 (1996).

How desire dies

MANY a woman complains about the short attention span of her lover. After orgasm he suddenly ceases to reiterate his love and passion; instead, he just turns over and goes to sleep. The speed of this orgasmic mental transformation demands explanation. The depletion of a sexual hormone would take far too long. The one thing that changes mood this fast is the sudden injection of a powerful drug. Heroin, nicotine and the adrenaline of our panic syndrome work in seconds. Clearly the male system is flooded at orgasm with some equally fast-acting passion-killing metabolite. Not until this has slowly cleared away can sexual interest return. Women have a different sexual chemistry, and can climax repeatedly.

And this, says Daedalus, explains the failure of the centuries-old search for a true male aphrodisiac. There is no such thing; there is only an anti-aphrodisiac, whose absence from the body allows sexual interest to develop. The brain must have receptors which turn off sexual desire when the compound binds to them.

Some pharmaceuticals can also deaden sexual desire, as an unwelcome side effect. They include antidepressants such as thioridazine and fluoxetine, beta-blockers such as atenolol, and even bromide — the British Army has been suspected of lacing soldiers' tea with bromide to dampen their sexual urges. Presumably these compounds or their metabolites bind more or less strongly to the brain's anti-aphrodisiac receptors.

So DREADCO's biochemists are seeking the male anti-aphrodisiac in donated blood samples. Daedalus suspects that it builds up in the male system with age. The blood of sexually obsessed teenagers should be almost free of it; that of sexually retired elderly gentlemen should be its richest natural source. It should soar dramatically in a volunteer seducer just after a conquest. Compounds whose concentration varies in the right way will be checked for molecular similarity to the sex-sabotaging drugs. Once identified, the anti-aphrodisiac will be synthesized for sale.

DREADCO's 'Passion-Killer' will abolish sexual desire in seconds. As a natural male product, it will be free of the feminizing side effects of previous hormonal anti-aphrodisiacs. Any man in compromising circumstances — a sexual criminal anxious not to re-offend, a diplomat cornered by some Mata Hari, or a politician striving to uphold family values — can simply 'pop' a pill and be released from temptation. Many a sexually reluctant woman will welcome it too, as a crafty additive to her husband's evening meal, or the candle-lit dinner of some unalluring lover.

David Jones

Bacterium genome sequence

SIR—Stephen Oliver in his Progress article champions a systematic approach to the discovery of gene function so as to make best use of the copious data being produced by large-scale genome sequencing projects¹. He gives cogent arguments for using yeast for this purpose, discusses the problem of functional redundancy and suggests ways of analysing open reading frames revealed by sequencing but having no matches with sequences in databases. Even open reading frames assigned on the basis of similarity to already-known genes can be problematic because of the paucity of functional information in the databases. Of relevance to this are our genomic studies of the photosynthetic bacterium *Rhodobacter capsulatus*, described here. These studies take advantage of this organism's unique properties, which include a simple system for determining the functions of sequenced regions.

R. capsulatus can be grown heterotrophically or phototrophically. It metabolizes a wide range of organic compounds and has two independent systems for fixing nitrogen. The genome consists of a single chromosome of 3.7 megabases and a 134-kilobase (kb) plasmid². A high-resolution physical map with more than 3,000 sites for four restriction enzymes contains 250 mapped genes. An overlapping set of 192 cosmids covers the genome completely. Most importantly, it harbours a defective transducing phage, called the gene transfer agent (GTA), which packages 4.6-kb DNA fragments representing all parts of the DNA present in a GTA-producing strain³. The GTA can be used to construct deletions efficiently and systematically, an important feature noted by Oliver for yeast.

For *Rhodobacter*, the fragment to be deleted can be replaced with a cartridge encoding resistance to kanamycin or gentamicin, or both (Fig. 1). This method was originally used to delete the entire *nifHDK* region⁴. We have exploited it more recently in attempts to delete whole-cosmid-sized regions of the chromosome. In these first experiments, we used 26 cosmids covering 400 kb of the *Rhodo-*

bacter chromosome. All 26 of the GTA lysates produced Km^r/Gm^r transductants, but with a wide range of morphologies and growth rates on complete medium. About half the cosmids yielded strains that had wild-type growth rates on complete medium. Figure 2 shows two of these deletions. The Southern blots indicate that, for cosmids 2D11 and 2D12, nearly all the chromosomal DNA contained in each cosmid is replaced by

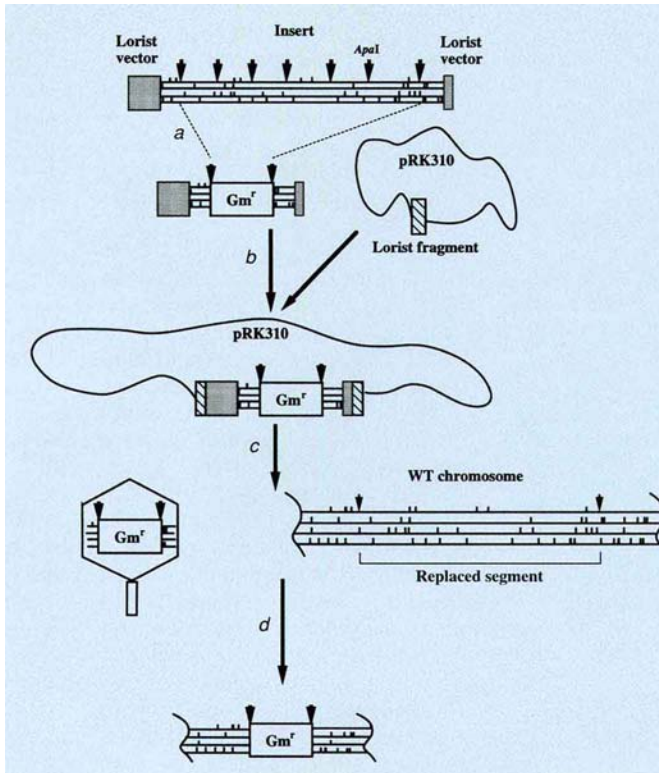


FIG. 1 Construction of deletion strains with the GTA. Cosmids containing 35–40-kb *R. capsulatus* DNA in the vector Lorist 6 are restricted with *Apal*, which cuts many times in *Rhodobacter* DNA but never in the vector (a). The cassette carrying resistance to both gentamicin and kanamycin is ligated into the cut cosmid and used to transform an *Escherichia coli* strain carrying pRK310, which in turn contains a fragment of the Lorist vector, allowing co-integrants to be formed (b). This construct is transferred to a GTA-producing strain of *Rhodobacter* by conjugation. Culture supernatants contain some GTA particles that have packaged the resistance cassette flanked by *Rhodobacter* DNA. Adding such supernatants to wild-type *Rhodobacter* (c) and plating on antibiotics gives transductants (usually) in which the target sequence is replaced by the resistance cassette (d). Selection on two antibiotics reduces the background of spontaneous resistant mutants.

the antibiotic-resistance cassette. To date, we have confirmed the complete deletion of 55 kb covered by three overlapping cosmids (2D11, 2D12 and 2E1) and of 25 kb each, covered by 2E11 and 2H3. Maps of these cosmids are shown in ref. 5.

Another group of transductants showed more complex hybridization patterns, to be expected where essential genes are located. In some cases, merozygotes were selected, in which the cassette was inserted in the target sequence but some of the target sequence was maintained. A blot showing this phenomenon for cosmid 1A3 is included in Fig. 2. The complete sequence of this cosmid has been determined (V.K., M.F. and R.H., manuscript in preparation).

Another aspect of analysis considered desirable by Oliver is the study of gene expression, preferably in a way that permits global detection of the consequences of mutation in a regulatory gene. This type of analysis has been demonstrated

for *R. capsulatus*, again exploiting the set of overlapping cosmids. The complete cosmid set was digested with *EcoRV* and the resulting fragments from each cosmid, after separation by electrophoresis, were transferred to a small number of blots. Effectively, the entire chromosome was displayed in 192 lanes in gene- or operon-sized pieces, according to their position on the chromosome. This high-resolution hybridization template was then used to detect, for example, the changes in transcription that accompany heat shock⁵. The same template can be used quantitatively

to detect changes in transcription in deletion mutants.

Over the past four months, we have completed 250 kb of sequence (GeneBank accession number U57682 for cosmids 143–147). The project is being carried out cosmid by cosmid, so progress can be monitored in terms of gene assignments as the work continues. So far, about half the open reading frames can be assigned with high probability. Another 10–15% contain a recognized functional domain but no clear indication of the full protein's function. The remaining 35–40% open reading frames have no matches in any public database. Sequencing of the *R. capsulatus* chromosome is being carried out without massive financial support. A grant from the US Department of Energy program in Basic Energy Sciences has funded all the mapping and cosmid library construction to date. Part of the sequencing is being done in Prague by V. Paces and C. Vlcek at the Academy

1. Oliver, S. G. *Nature* **379**, 597–600 (1996).
2. Fonstein, M. & Haselkorn, R. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2522–2526 (1993).
3. Yen, H. C. et al. *J. molec. Biol.* **131**, 157–168 (1979).
4. Scolnik, P. A. & Haselkorn, R. *Nature* **307**, 289–292 (1984).
5. Fonstein, M. et al. *EMBO J.* **14**, 1827–1841 (1995).

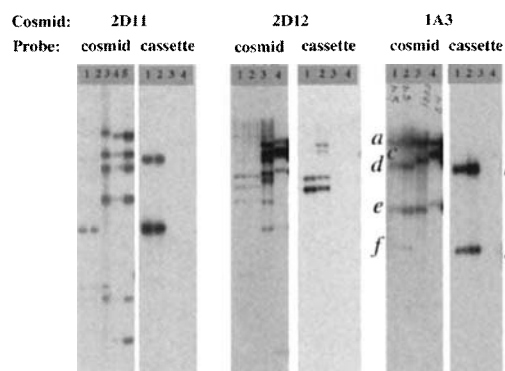


FIG. 2 GTA-mediated deletions in *R. capsulatus*. Each panel contains the following DNA samples restricted with *EcoRV*: lanes 1, 2, total DNA from two independent transductants; lane 3, total DNA from wild-type *R. capsulatus*; lane 4 (and 5 in the first panel), cosmid DNA. For each pair of panels, the hybridization probes were the total cosmid (left) or the antibiotic-resistance cassette (right), which contains a single *EcoRV* site. Discrepancies between fragment sizes in lanes 3 and 4 are due to vector sequences attached to *Rhodobacter* DNA in lane 4. For cosmids 2D11 and 2D12, most of the DNA present in the cosmids has been deleted from the transductants. For cosmid 1A3, fragments *a* (20 kb) and *e* (2 kb) cannot be deleted. Fragment *c* (11.5 kb) is disrupted by the cassette in the transductants, yielding fragments *d* and *f*; these fragments correspond in the right-hand panel to fragments *b* and *g*, which hybridize with the cassette probe, confirming the location of the disruption. The sum of the sizes of fragments *d* and *f* is 11–12 kb. Because the cassette size is 2 kb, the deletion in these transductants is only 1–2 kb. Fragment *c* is close to one end of cosmid 1A3. We believe that the recombination event creating these deletions began in the region of fragment *c* flanking the cassette and that the second crossover occurred in a region of partial homology located nearby, rather than at the far end of the cosmid where homology with the GTA-carried DNA is perfect.

of Sciences of the Czech Republic, with a grant from the NSF for international collaboration.

**Vivek Kumar, Michael Fonstein
Robert Haselkorn**
Department of Molecular Genetics
& Cell Biology,
The University of Chicago,
Chicago, Illinois 60637, USA

OLIVER REPLIES — Kumar and colleagues demonstrate that techniques now exist in a wide range of organisms which will permit the systematic analysis of gene function. Such an analysis is most meaningful in organisms such as *Saccharomyces cerevisiae*, whose genome has been completely sequenced, and it is encouraging to learn that an international effort to sequence the *Rhodobacter capsulatus* genome has begun.

The techniques outlined by Kumar *et al.* for *R. capsulatus* have quite exact analogues in other systems. In *S. cerevisiae*,

the complete genome sequence, together with the organism's high efficiency and accuracy of mitotic recombination, is being exploited in a gene replacement strategy involving the polymerase chain reaction (PCR)¹, and resulting in the efficient deletion of individual open reading frames (ORFs).

Following the successes of our European Yeast Genome Sequencing Network², we have formed another scientific network (EUROFAN) committed to elucidating the function of novel yeast genes. In EUROFAN, we are exploiting a gene-by-gene deletion strategy, using PCR-mediated replacement and (as with *R. capsulatus*) *Gm^r* as a replacement marker^{3,4}. Some laboratories are also performing more extensive chromosomal deletions using a combinatorial approach⁵, which we term "mass murder".

A parallel European network is working on functional analysis in the Gram-positive bacterium *Bacillus subtilis*. This group is exploiting the operon organization common for bacterial genes by using a single plasmid insertion event to (simultaneously) disrupt an ORF, create a *lacZ* fusion with the upstream ORF, and place the downstream ORF under the control of a regulatable promoter (S. D. Ehrlich, personal communication).

As with more conventional approaches to defining gene function, different model organisms will be used for the systematic analysis of particular biological systems. Thus, *R. capsulatus* may be used for an exhaustive genetic analysis of nitrogen fixation or photosynthesis, while *S. cerevisiae* may be used to study functions that are peculiar to eukaryotes and thus not accessible to experimentation with a bacterium.

Within a given organism, it will be important to construct strains which have an improved potential for elucidating the function of novel genes in particular areas of biological activity⁶. More generally, there is a need to develop global approaches to the analysis of gene function, such that data from these model organisms can be exploited rapidly for analysing species that are less genetically malleable, but which have great biological, medical, agricultural or industrial interest. In particular, it will be important to develop strategies to permit the 'functional mapping' of the genomes of such species onto those of the model organisms.

S. G. Oliver
Department of Biochemistry
& Applied Molecular Biology,
UMIST, PO Box 88,
Manchester M60 1QD, UK

Contamination of drinking water

SIR — Until now, the availability of organic carbon has been considered the key factor controlling microbial regrowth in drinking-water networks^{1,2}. This availability is considered to be a potentially serious problem, especially in boreal regions (northern Europe, Russia and North America) where surface water, and sometimes ground waters, contain high amounts of organic matter³. Natural organic matter acts as a substrate for microbial growth and, when water is used for public supply, disinfectants such as chlorine, hypochlorite and ozone are used. In addition to killing microorganisms, these oxidizing agents break down large organic compounds to smaller ones^{4,5}. These compounds can be easily decomposed, further increasing microbial growth in the distribution system and thus impairing water quality. Because of this, considerable effort and resources have been deployed to remove the organic contaminants from drinking water. In the United States, for example, current annual expenditure exceeds 5 billion dollars⁶.

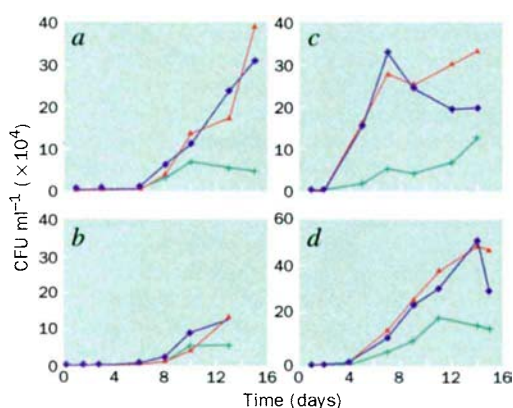
Finnish drinking waters contain high concentrations of organic carbon (75–640 µg C l⁻¹) easily available to microorganisms. But because the available organic carbon correlates negatively with microbial growth, we investigated the possibility that other factors regulate this growth. In experiments with ground and surface waters, addition of phosphate alone to water increases microbial growth (the number of culturable bacteria) to the same extent as does a mixture of added inorganic nutrients (see figure).

The concentration of phosphate needed to enhance microbial growth is extremely low. When we added phosphate ranging from zero to 50 µg l⁻¹ to the processed drinking water supplying five waterworks (three with surface water and two with ground water), we found a hyperbolic relationship between [PO₄-P] and microbial growth, with a half-saturation value of only 2–3 µg PO₄-P l⁻¹. Investigation of samples from four distribution

- Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. *Nucleic Acids Res.* **21**, 3329–3330 (1993).
- Vassarotti, A. *et al.* *J. Biotechnol.* **41**, 131–137 (1995).
- Wach, A., Brachat, A., Pohlmann, R. & Philippsen, P. *Yeast* **10**, 1793–1808 (1994).
- Wach, A. *Yeast* **12**, 259–265 (1996).
- Fairhead, C., Llorente, B., Denis, F., Soler, M. & Dujon, B. *Yeast* (in the press).
- Oliver, S. G. *Nature* **379**, 597–600 (1996).

- LeChevallier, M. W., Babcock, T. M. & Lee, R. G. *Appl. Environ. Microbiol.* **53**, 2714–2724 (1987).
- van der Kooij, D. *J. Am. Wat. Works Ass.* **84**, 57–66 (1992).
- Cronan, C. S. in *Organic Acids in Aquatic Ecosystems* (eds Perdue, E. M. & Gjessing, E. T.) 245–260 (Wiley, Chichester, 1990).
- van der Kooij, D., Hijnen, W. A. M. & Kruitthof, J. C. *Ozone Sci. Engng* **11**, 297–311 (1989).
- Daniel, P., Meyerhofer, P., Zafer, M. & Rice, E. in *Proc. 11th Ozone World Congr. Ozone in Water and Wastewater Treatment Vol. 2* (Int. Ozone Ass., Stanford, 1993).
- van den Leeden, F., Troise, F. L. & Todd, D. K. *The Water Encyclopedia* 2nd edn (Lewis, London, 1990).
- American Public Health Association, American Water Works Association & Water Environmental Federation *Standard Methods for the Examination of Water and Waste Water* (eds Greenberg, A. E., Clesceri, L. S. & Eaton, A. D.) (Am. Publ. Hlth Ass., Baltimore, 1992).

Growth of heterotrophic bacteria in different Finnish drinking waters. *a, b*, Two waterworks using surface water; *c, d*, two waterworks using ground water. We incubated water samples taken from the waters produced in these waterworks at 15 °C in the dark. We monitored bacterial growth in samples by counting bacteria (colony-forming units (CFU) per ml) every second day by a spread-plate technique with R2A-agar⁷. We performed tests once at each site. Tests included untreated samples (green), samples with addition of phosphorus (red) (50 µg P l⁻¹) and samples amended with a mixture of inorganic nutrients (blue) ((NH₄)₂SO₄, KH₂PO₄, MgSO₄ × 7 H₂O, CaCl₂ × 2H₂O, NaCl). The final concentrations of added inorganic ions were: 970 µg NH₄-N l⁻¹; 50 µg PO₄-P l⁻¹; 60 µg K l⁻¹; 10 µg Mg l⁻¹; and 30 µg Ca l⁻¹.



networks, whose water came from two ground and two surface waterworks, showed a similar large stimulation (>ten-fold) of bacterial growth when small amounts of phosphate were added.

Because phosphorus is an important factor regulating microbial growth in drinking water, we suggest that water management organizations should consider phosphorus removal. The development of stringent phosphorus removal

measures could lead to more cost-effective methods for restricting microbial growth and hence improving the quality of drinking water.

Ilkka T. Miettinen
Terttu Vartiainen
Pertti J. Martikainen

Division of Environmental Health,
National Public Health Institute,
PO Box 95,
70701 Kuopio, Finland

Infanticidal male eresid spiders

SIR — The reproductive interests of males and females rarely coincide, being most polarized when the parental effort of each sex is highly asymmetrical, for example when females alone care for the brood. Males may prevail in sexual conflicts even to the extent of forcing their interests on females¹. Probably, the costliest male strategy to a brood-caring female is infanticide, where a male kills her offspring. So far, male infanticide is known only in mammals^{2,3} and birds⁴. We have found male infanticide (oocide) in a spider species in which the females exhibit extreme maternal care: *Stegodyphus lineatus* (Eresidae) females provide suicidal care to the matrophagous hatchlings of her

single clutch. Thus, once having produced eggs, a female does not benefit from further matings. But a male that fights with a female and wins will remove the egg sac, forcing the female to replace her clutch.

Spiders have provided dramatic examples of sexual conflict in which the male loses by being cannibalized after, during or even before mating^{5,6}. Males are frequently at a disadvantage because they are generally smaller than females^{5,7}. However, the eresid spider *S. lineatus* provides an example of a sexual conflict in which females lose despite their size advantage. Female *S. lineatus* usually produce only one egg sac, although they can lay replacement clutches if it is removed (J.M.S. and Y.L., manuscript in preparation). The young cannot leave the egg sac on their own; even after their mother releases them from the cocoon they depend on her completely for food and protection during the two weeks after hatching⁸. The female always dies in the course of brood care because the young kill and consume her two to three weeks after hatching⁸. Thus, females make a terminal investment of all their resources in caring for a single clutch, whereas males contribute no parental care (see ref. 9).

We studied a natural population of 278 female spiders with egg sacs on the Avedat plateau in the central Negev desert, Israel, of which 63 (22.7%) lost their first clutch. Males were responsible for the loss of the egg sacs in 33% of these cases. In *S. lineatus*,

the mating and egg-laying seasons overlap, such that late-maturing males are still searching for females when up to 50% of the females already have egg sacs (Fig. 1). Loss of the egg sac is costly to a female; therefore, it is not surprising that females defend them aggressively against males.

In field tests, we placed males of known size and reproductive history at the edges of webs of females guarding egg sacs. Females succeeded in chasing males off their webs in 28 (51%) of 55 experimental contests. Males sustained injuries in 3 (5.5%) instances. A male that succeeded in entering the tube-like nest of a female guarding her egg sac first used his chelicerae — appendages modified as pincer-like jaws — to remove the threads that attach the egg sac to the inner wall of the nest. Then, he manoeuvred the egg sac to the nest entrance and eventually dropped it to the ground. Once the egg sac was removed, the female could not retrieve it.

What do males gain from infanticidal behaviour? We determined the mating success of males, individually marked at maturation, by checking all female nests for male presence (Fig. 2). Using presence in females' nests as an indication of mating success, we estimated that a male can expect to encounter and mate with an average of 1.2 females. Thus, if a searching male encounters a female with an egg sac and does not mate with her, he is likely to lose most of his expected mating success.

In entelegyne spiders such as the Eresidae, the female's sperm-storage organ (spermatheca) has separate ducts for insemination and fertilization¹⁰. Spermathecal morphology suggests that the sperm of the first male to mate with a female will have priority in fertilizing her eggs¹⁰. Furthermore, clutches of *S. lineatus* are small (40–140 eggs); sperm depletion is unlikely because we have observed that females produce viable second egg sacs without additional mating. In such a case of first-male sperm priority and no sperm depletion, an infanticidal male could expect little fertilization success.

We examined the sperm priority pattern

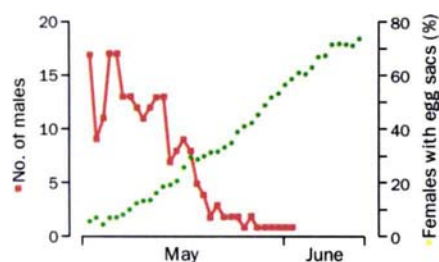


FIG. 1 Overlap of mating and egg-laying seasons; we show number of active males (■) and the cumulative proportion of females with eggs (●) from a total population of 320 adult females. Data were obtained in 1993 from a population inhabiting a ~2-km-long dry river bed in the Negev desert, Israel.

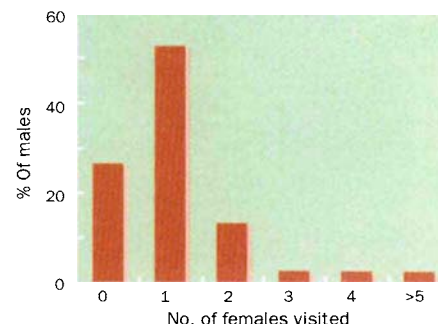


FIG. 2 Number of female nests visited by males which had been marked for individual recognition before they left their nests to search for females ($n = 38$). Most males were found in one or in two female nests. Predation risk for males while searching for a mate is probably very high.

in *S. lineatus* using a sterile-male technique, in which irradiated sperm competes with normal sperm for 'fertilizing' eggs, but after 'fertilization' no embryo develops. By irradiating males with 15 krad γ -radiation from a ^{60}Co γ -emitter (Department of Nuclear Engineering, Ben Gurion University), sperm were made sterile without affecting male performance. Females were paired with two males in all combinations of normal (N) and sterile (S) males: NN ($n = 9+13$ from the previous year), SS ($n = 10$), SN ($n = 14$) and NS ($n = 10$). None of the eggs in the SS matings hatched. We corrected the hatching success of clutches from SN and NS matings by the proportion of unhatched eggs in the NN controls (13.6%). There was no first-male priority. Indeed, the calculated probabilities of fertilization by first (P1) and second (P2) males to mate suggest a complete mixing of sperm: $P1(\text{SN}) = 0.417$, $P1(\text{NS}) = 0.607$, $P2(\text{SN}) = 0.583$ and $P2(\text{NS}) = 0.393$. Sterile sperm had a slightly lower fertilization probability than normal sperm, but this was the case regardless of whether the first or the second male was sterile.

Complete sperm mixing in the spermatheca means that a male can expect a similar fertilization success in matings with females with or without eggs. The adaptive value of male infanticide in these spiders is that males increase their reproductive success at the expense of males that matured earlier in the season, although at a cost to the females. For males, infanticide is always advantageous because their reproductive success increases with the number of matings, which is perhaps the only option for late-maturing males. But there is a risk. For females that lose their clutch to an infanticidal male, there are costs of lower survival and reduced fecundity (our unpublished data). Also, the offspring of females who raise a replacement clutch disperse late in the season and have less time to grow before the abundance of flying insects decreases. Smaller juveniles have a lower probability of survival¹¹. This may

also be a cost for the male which is expressed in the reduced fitness of his offspring. However, a male that matures late in the season will have late-dispersing offspring regardless of whether he mates with a female with eggs or a virgin female.

Infanticidal males of the spider *S. lineatus* gain matings, but decrease female survival and fecundity. This male behaviour can persist only if the costs to the male's own offspring are smaller than the direct benefits to him¹². Male infanticide has not been observed previously in spiders, nor

to our knowledge in other invertebrates. However, sexually selected infanticide may be more common than previously recognized. We predict that male infanticidal strategies will be found in more invertebrate species that exhibit extreme maternal brood care.

Jutta M. Schneider*, Yael Lubin
Mitrani Center of Desert Ecology,
Jacob Blaustein Institute of
Desert Research,
Ben Gurion University of the Negev,
84990 Sede Boker, Israel

Protein-RNA molecular recognition

SIR—The non-canonical G•U wobble pair is an important secondary structural feature of several RNA helices. Although connected by two hydrogen bonds, G•U differs in its geometry from a Watson-Crick base pair. It participates in both protein-RNA^{1,2} and RNA-RNA³ interactions. The G•U pair found in the acceptor helix of *Escherichia coli* alanyl-transfer RNA (tRNA^{Ala}) has recently been found to contribute *in vivo* to the interaction of this RNA with the enzyme alanyl-tRNA synthetase, by introducing a helical distortion¹. This result challenges previous *in vitro* studies which concluded that the role of the G•U base pair in tRNA^{Ala} is to allow alanyl-tRNA synthetase to recognize directly the 2-amino and 2'-hydroxyl groups in the minor groove of the A-form RNA helix². But, based on recent studies into the recognition of tRNA^{Gln} by *E. coli* glutamyl-tRNA synthetase⁴, we suggest that this discrepancy can be reconciled by considering the diverse *in vivo* and *in vitro* conditions in these two studies.

The *in vitro* studies, which involved quantification of the alanylation of RNA molecules with various substitutions of G•U, were performed at a subsaturating concentration of alanine (22 μM ; Michaelis constant, $K_m = 240 \mu\text{M}$; ref. 5) due to inherent limitations of the assay system. Subsequent analysis of these experiments assumed that the use of diverse RNA substrates would not affect the kinetic parameters of alanyl-tRNA synthetase for alanine. But this critical assumption may not be valid, as it has now been shown for the glutamyl- and tryptophanyl-tRNA synthetases that the mutation of certain nucleotides in the tRNA leads to substantial changes in the kinetic parameters of these synthetases with respect, not only to the tRNA, but also to the amino-acid substrate⁶. The replacement of G3•C70 by G•U in the acceptor helix of tRNA^{Gln}, for example, leads to an almost threefold increase in the K_m for glutamine but no significant change in the catalytic constant, k_{cat} . If a comparable situation were to exist during the recognition of tRNA^{Ala} variants by alanyl-tRNA synthetase, use of a uniform subsaturating

concentration of alanine would lead to non-uniform underestimation of k_{cat} , and consequently to the misassignment of energetic contributions to binding.

In contrast, the *in vivo* assay of this protein-RNA interaction is less susceptible to such errors, because the concentration of alanine in the amino-acid pool of *E. coli* is almost an order of magnitude higher than used *in vitro* (168 μM ; ref. 7). Nevertheless, the *in vivo* assay also suffers from a limitation in that the concentration of alanyl-tRNA synthetase may depend on several metabolic parameters, leading to the possibility that the observed differences in tRNA alanylation may be due to changes in the level of the enzyme itself, rather than changes in the structure of tRNA^{Ala}. Add to this the possibility that alanyl-tRNA synthetase undergoes tRNA-dependent changes in its kinetic parameters for alanine, and what initially seem to be contradictory results on the recognition of G•U by alanyl-tRNA synthetase can, in fact, be readily reconciled. In this way, a nucleotide replacement that substantially reduces the level of tRNA charging under certain conditions *in vitro* may be effectively compensated for *in vivo* by increased concentrations of both amino acid and synthetase, particularly when the gene encoding alanyl-tRNA synthetase is overexpressed⁴.

Although the problem in interpreting these results raises the perennial concern of the validity of comparing *in vitro* and *in vivo* approaches to studying molecular recognition, it shows that in this case the question of direct or indirect recognition will finally be resolved only by determining the crystallographic structure of tRNA^{Ala} complexed with alanyl-tRNA synthetase.

Michael Ibba, Dieter Söll
Department of Molecular Biophysics
and Biochemistry,
Yale University, 266 Whitney Avenue,
New Haven, Connecticut 06520-8114, USA

- Clutton-Brock, T. H. & Parker, G. A. *Anim. Behav.* **49**, 1345–1365 (1994).
- Hrdy, S. B. *Ethol. Sociobiol.* **1**, 13–40 (1979).
- Bertram, B. in *Growing Points in Ethology* (eds Bateson, P. G. & Hinde, R. A.) 281–301 (Cambridge Univ. Press, 1976).
- Rohwer, S. in *Current Ornithology* Vol. 3 (ed. Johnston, R. F.) 353–391 (Plenum, New York, 1976).
- Elgar, M. A. & Crespi, B. J. (eds) in *Cannibalism: Ecology and Evolution among Diverse Taxa* 1–12 (Oxford Univ. Press, 1992).
- Andrade, M. *Science* **271**, 70–72 (1996).
- Vollrath, F. & Parker, G. A. *Nature* **360**, 156–159 (1992).
- Schneider, J. M. *Insect Soc.* **42**, 237–248 (1995).
- Evans, T. et al. *Nature* **376**, 299 (1995).
- Austad, S. N. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 223–249 (Academic, London, 1984).
- Schneider, J. M. thesis, Univ. München (1992).
- Parker, G. A. in *Sexual Selection and Reproductive Competition in Insects* (eds Blum, M. S. & Blum, N. A.) 123–166 (Academic, New York, 1979).

*Also at: Max-Planck-Institut für Verhaltensphysiologie, Seewiesen, D-82319 Starnberg, Germany.

- Gabriel, K. et al. *Science* **271**, 195–197 (1996).
- Musier-Forsyth, K. & Schimmel, P. *Nature* **357**, 513–515 (1992).
- Strobel, S. A. & Cech, T. R. *Science* **267**, 675–679 (1995).
- Hong, K.-W. et al. *EMBO J.* **15**, 1983–1991 (1996).
- Shi, J.-P. et al. *Biochemistry* **33**, 5312–5318 (1994).
- Ibba, M. et al. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Raunio, R. & Rosenqvist, H. *Acta chem. scand.* **24**, 2737 (1970).

Analysis of a mind

Philip Kitcher

Bertrand Russell: The Spirit of Solitude. By Ray Monk. Cape: 1996. Pp. 680. £25.

FOR young Britons of my generation with intellectual interests and left-wing sympathies who came of age in the 1960s, Bertrand Russell was an inspiration, combining lucidity and precision with political passion, as if a mind of exceptional power had fused with the body of an Old Testament prophet, sent to call the world to account. His writings and speeches on nuclear disarmament drew us to his political and philosophical books, first to the more popular works on social questions, later to the metaphysics and epistemology. We glimpsed, looming behind them, the peaks of his achievement, the works on logic, remote and difficult. For some of us, one of the rewards of turning to academic philosophy came in ascending those forbidding slopes and appreciating something of what Russell contributed to the history of Western thought.

In 1967, he published the first volume of his *Autobiography*, and our vision of him changed. Russell's own narrative of his early life revealed unsuspected vulnerability, and his apparently uncompromising account of his personal relationships suggested that an unhappy childhood had produced a young adult capable of considerable callousness and cruelty. His description of the end of his first marriage records how he threatened Alys Russell that he would commit suicide if she used the name of his lover, Lady Ottoline Morrell, in divorce proceedings. Russell continues:

I meant this, and she saw that I did. Thereupon her rage became unbearable. After she had stormed for some hours, I gave a lesson in Locke's philosophy to her niece, Karin Costelloe, who was about to take her Tripos. I then rode away on my bicycle, and with that my first marriage came to an end. I did not see Alys again till 1950, when we met as friendly acquaintances.

Since Classical Antiquity, philosophers have been supposed to be abstracted from worldly concerns, but the kind of ruthless detachment portrayed here hardly appears to conform to any historical ideal.

Given Russell's own seeming willingness to present himself naked to the world, we might wonder what has been

left for subsequent biographers. Ray Monk's fascinating study reveals that there is much more to learn about his subject. Tracing Russell's career from the oppressive atmosphere of Pembroke Lodge, heavy with the ghosts of the Victorian past, to his marriage to Dora Black and the birth of his first child, Monk skilfully weaves together Russell's intellectual accomplishments, his dedication to political causes and his personal

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASON
REASONS

Russell: confessed to being driven by "the longing for love, the search for knowledge, and unbearable pity for the suffering of mankind".

life. At the centre of the book is the enormous enigma posed by the contrast between the public and private personae, the one scrupulously honest, courageous and compassionate, the other cold, egoistic, often cowardly and sometimes duplicitous. Monk brilliantly tells the story of a series of quests, pursued by a young man thirsty for knowledge, anxious to distinguish himself, intensely affected by the perception of injustice and, above all, yearning desperately for love.

Monk's thorough research and his talent for analysis enable him to correct details, often important ones, that Russell did not recall, and to probe hidden motives to which Russell was, quite understandably, blind. Russell's account of the end of his first marriage is chilling, but it has its own kind of purity — the reality was messier. Before the final break, Russell was prepared to bribe Alys into refraining from making a public fuss by allowing her "very small amounts of

my society"; and the detached figure cycling into the future had to return occasionally for some extremely fraught negotiations.

Monk's presentations of Russell's philosophical views are always clear and accurate, but, despite his attempts to make the intellectual ventures central to his story, they are typically upstaged by the personal turmoil. It will surely be hard for the nonspecialist to obtain from this book any detailed picture of Russell's philosophical development, or even a sense of the urgency Russell felt about finding a solution to the paradoxes of set theory, nor can Monk convey the great difficulty of the logical problems. This is not, in my judgement, a remediable flaw: rich biography and intricate philosophical reconstruction are hardly compatible bedfellows. (For an excellent study of Russell's intellectual development and of the significance of his early work, see Peter Hylton's *Russell, Idealism, and the Emergence of Analytic Philosophy*; Oxford University Press, 1990.)

It would be wrong, however, to think that Monk sheds no new light on Russell's philosophical work. His compelling description of the relationship between Russell and Wittgenstein, conjuring a vision of the insecurity felt by the senior scholar confronted by a young man whose disturbing genius was patent, helps us to understand how Russell could have been directed along the philosophical paths he pursued after the publication of *Principia Mathematica*. Moreover, in watching Russell measure himself against his "young Austrian", we can appreciate the attraction for him of excursions into politics, and, later in life, into less esoteric branches of philosophy.

Yet the most significant contribution of the biography is its connection of the philosophical work and the political causes to the psychologically complex individual who produced them. Russell began his *Autobiography* with a prologue in which he saw himself driven by three passions: "the longing for love, the search for knowledge, and unbearable pity for the suffering of mankind". Monk makes us feel these passions pulsating through the early years of Russell's life. On many occasions, the first is repellent in its power and in its direction, for the longing for love was primarily a compulsion to receive love, so that Russell's letters to the women whose lives he touched — and sometimes left in ruins — are dominated by his own personal needs,

Hulton Deutsch

from them, from his politics and from his philosophical work, and the principal voice of the letters is a great yammering "I", full of urgent demands that some of those he addressed felt to be irresistible. If we feel pity for those we see as his victims, Monk makes us spare pity for Russell too, himself a victim, while allowing us still to retain our admiration for a man who felt so keenly two other great passions, and whose philosophical and political achievements remain as monuments to an extraordinary life. □

Philip Kitcher is in the Department of Philosophy, University of California at San Diego, La Jolla, California 92093, USA.

Beyond the Garden of Eden

Jean-Jacques Hublin

The Neandertal Enigma: Solving the Mystery of Modern Human Origins. By James Shreeve. Viking: 1996. Pp. 369. £18.

African Exodus: The Origins of Modern Humanity. By Chris Stringer and Robin McKie. Cape: 1996. Pp. 267. £18.99.

Humans Before Humanity. By Robert Foley. Blackwell: 1995. Pp. 238. £25, \$29.95.

In 1856, when a primitive human skeleton was discovered in Neanderthal, a valley near Düsseldorf, Germany, it was thought to have belonged to a pathological idiot, a lost Cossack and, by some scientists, a member of an extinct human race. Are the Neanderthals still a mystery today, as James Shreeve suggests? If so, the mystery is about to be resolved. After 140 years of investigation, we know more about these predecessors of modern humans in Europe than we do about any other group of fossil hominids. Their territory is certainly the area where human evolution and environmental change over the past 500,000 years have been best documented. What is more, Europe shows a clear-cut picture of the relationships between these archaic populations and the later modern invaders who appeared 40,000 years ago. The counterpoint to the rise and fall of the Neanderthals is the origin and evolution of modern humans, and the task is to propose an up-to-date interpretation of this 'score for two voices'.

Shreeve's well-written book provides an outsider's view that is blessedly accessible to the nonspecialist. Its spicy dramatization and lyricism are the mark of the professional science writer. But its primary attraction lies in its construction. More than an enquiry into a scientific

debate, it observes in minute detail the scientists involved. Shreeve has travelled the world meeting many of the strange people who devote their lives to the study of stone artefacts, human bones and, most important of all, opposing evolutionary models, and he provides us with an enjoyable gallery of portraits. His journey starts in a Parisian café, and then continues in museums, laboratories, homes, tents and caves. We see science in action, particularly the mixture of perceptiveness and intelligence — as well as irrationality — characterizing the main players. And we are taken behind the scenes of scientific meetings, where bitter academic wars are not just academic.

roots of the multiregional model and breaking it down.

In a section entitled "A personal history", Stringer reveals his academic agenda. Like many other scholars, he was deeply influenced in building a so-called 'replacement' model (in which archaic humans were subsumed by the spread of modern human populations) by his study of the European palaeoanthropological evidence, which is dominated by the Neanderthals. The second pillar of the 'Out of Africa' model is then revealed, and the reader will find a selected review of the genetic studies demonstrating the great homogeneity of living populations and suggesting their likely African



Alan Walker attending to the skull of Nariokotome Boy', the most complete skeleton of *Homo erectus* found so far. In *The Wisdom of the Bones: In Search of Human Origins*, written with his wife, Walker gives a vivid account of his discovery and its implications for the evolution of modern humans. Weidenfeld and Nicolson, £18.99.

With his relentless quest for the truth and urge to tell a good story, Shreeve could be reproached for his faith in 'Eve'. The 'Out of Africa' model, which holds that all modern humans descend from a limited group of African hominids (including 'Eve') less than 200,000 years old, is preferred to the 'multiregional model', which argues for the emergence of modern races throughout the Old World from local archaic populations. Many readers will simply see here the victory of today's orthodoxy in human palaeoanthropology.

Such is the view of the anthropologist Chris Stringer, who has teamed up with Robin McKie, a science journalist, to write *African Exodus*. In this case, the Neanderthals, although unavoidable, are not the focus of their concern. After reviewing millions of years of human evolution and important concepts about species, races and adaptation and so on, the Stringer and McKie reach the heart of their project: explaining the historical

origins. One could argue that the debatable assumptions lying behind some results are not always exposed, but who better to defend the model than its most ardent advocate?

The defence is skilful and enthusiastic. Topics that often appear obscure and impenetrable to the public are made crystal-clear. Even after years of having read about this controversy, I could hardly put the book down. Another attractive feature of the book is its attempt, in the light of the 'Out of Africa' model, to propose a global interpretation of human development. The milestones of the exodus are explored in each geographical area, but also their consequences for the interpretation of the physical and mental traits of living humans.

Robert Foley's book is broader in scope, encompassing the evolution of all the hominids. In contrast to most of the books on this popular subject, Foley does not provide us with a fresco showing the ascent from ape to angel, or at least to a

Promethean hero. His goal is not to give all the historical details of the epic; rather, it is to explore the questions surrounding the evolution, on a late branch of the primate phylogenetic tree, of this apparently improbable freak group once thought to be of divine origin: a bipedal sweaty ape afflicted with a costly large brain. What are human beings? Why Africa? When did we become humans? Why did humans evolve?

Foley offers clear and often original answers to crucial questions. His perspective is definitely Darwinian and sometimes (refreshingly) non-politically correct. Drawing on biology, palaeontology, ecology and ethology, he explains how small causes scattered over vast geological times have produced great effects. A number of conventional ideas are usefully broken and the result is really fascinating. Anyone interested in evolution must read this book.

Our origins are definitely tropical. And to varying degrees, virtually all palaeoanthropologists now support the notion of humans having an African origin and evolutionary continuity on that continent. Africa was until recently the core demographic and selective environment of hominids. Strongly opposing models mainly concern areas that remained marginal for most of hominid evolution. Indeed, human palaeontology was born outside the vast garden of Eden, in Europe where archaic humans evolved away from the mainstream and were finally submerged. But other marginal areas are either *terra incognita* or do not fit so easily into a simplistic 'replacement' model.

In fact, the Far Eastern evidence is relatively little commented on in these three books. How many 'Out of Africa' events do we need to interpret the Asian archaic hominids? How much parallel evolution can we see in these remote populations? Did they reach the status of a distinct species of hominid? If not, could there have been gene flow between different waves of human populations invading Asia? Enigmas still persist inside and outside the Garden of Eden. □

Jean-Jacques Hublin is at the CNRS Laboratoire d'Anthropologie, Musée de l'Homme, Place du Trocadero, 75116 Paris, France.

Classic reissue

Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought by George C. Williams. This hugely influential book, first published in 1966, combines original insight with a hard-headed defence of Darwinian selection. The author, one of the leading US evolutionary biologist, provides a new introduction.

Now that's what you call chamber music

MANY readers will have heard a version of the curious auditory illusion, deliberately constructed, whereby a sequence of musical notes seems continually to ascend in pitch. It is, in a loose sense, an aural analogue of Escher's two-dimensional lithograph in which monks endlessly ascend (and others endlessly descend) the stairs around the top of their illusory monastery. The aural illusion has various relatives, all of which I find amusing. But music they are not.

By contrast, the 'compositions' by Zach Davids on the new compact disc **Heartsongs: Musical Mappings of the Heartbeat** are definitely musical. Yet they are derived, in a direct way, from

time series of changes in heart rate. Our heartbeats in fact have a great deal of subtle variability from one beat to the next, caused by the normal workings of the involuntary nervous system. In the authors' words: "the normal heartbeat, therefore, does not follow a metronomic or march-like beat — surprisingly, it has a dance-like plasticity and variability".

The next step was to convert the time series of intervals between heartbeats into integers on a scale of 1 to 18. (This process requires smoothing: 330 integers were obtained for each dataset of 100,000 heartbeats; that is, the average was taken over successive runs of 300 beats.) The last step was to map

the numbers onto a musical scale, specifically the diatonic musical scale.

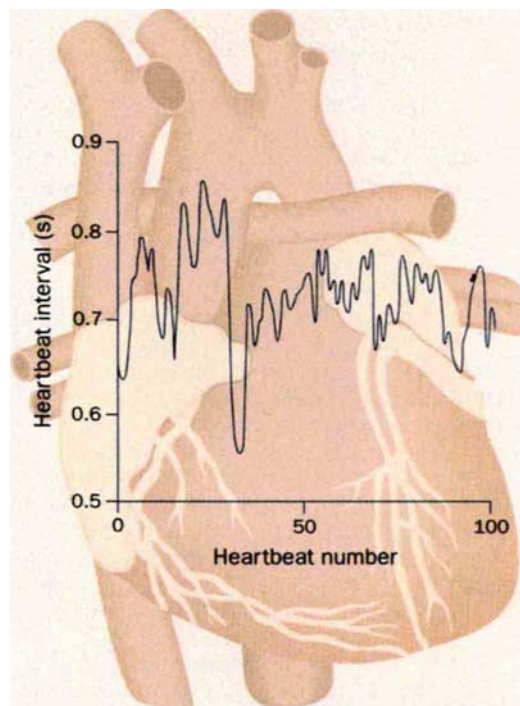
In this way, the sequences of integers generated the melody. Changes in pitch were proportional to the changes in heart rate. So much for the mechanics. The composer (Davids) was then "free to choose the rhythm and harmonic accompaniment for each melody", say the authors. "However, the underlying melodic line for each piece remains true to the original heartbeat time series." This leaves me in a bit of doubt about exactly how much of what one hears is undiluted 'heartsong'.

What is clear, however, is that the result is musically pleasing, at least to my unsophisticated ear. And, like Goldberger and colleagues, I think this raises interesting questions.

We could equally have ended up with boring sameness, or even dissonant jangle. The authors speculate that musical composition may involve, to some degree, "the recreation by the mind of the body's own naturally complex rhythms and frequencies. Perhaps what the ear and the brain perceive as pleasing or interesting are variations in pitch that resonate with or replicate the body's own complex (fractal) variability and scaling."

This is probably all a bit overheated, but the fact remains that this music — largely mechanically produced — is intriguing to listen to. Buy the CD and form your own view. □

Robert M. May is Chief Scientific Adviser to the UK government, on leave from the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.



Heartbeat fluctuations of a healthy person

digital tape recordings of actual heartbeats from 15 people at Harvard Medical School's Beth Israel Hospital.

The CD is published by Ivory Moon Recordings (22 Rutgers Road, Wellesley, Massachusetts 02181, USA; CD 7-88897-68002-3) and arises from a project developed as part of a major exhibit at the Boston Museum of Science (which opened in December last year) entitled "The Dance of Chance: Growing Order out of Randomness". It comes with explanatory notes written by Ary Goldberger and Zach Golberger (Davids's nonprofessional name), C.-K. Peng and Paul Trunfio, who are all at Harvard Medical School and Boston University.

Beginning with electrocardiogram data, the authors measured the precise intervals between heartbeats to obtain a

Damage limitation

Stuart Sutherland

Classic Cases in Neuropsychology.

Edited by Chris Code, Claus-W. Wallesch, Yves Joanette and Andre Roch Lecours. Lawrence Erlbaum: 1996. Pp. 385. £34.95.

PSYCHOLOGY, like shirts and skirts, is subject to the vagaries of fashion. Currently in vogue, particularly in the United Kingdom, is the examination of the effects of brain damage, whether caused by surgery, tumours or strokes, a discipline loosely known as neuropsychology (although that term can be applied to other approaches to brain function). The founder of the method was effectively Paul Broca, who in 1861 concluded that an area of the left frontal lobe produced motor aphasia — the inability to talk, with other motor functions and the capacity to understand speech left reasonably intact. Progress has been aided recently by the development of sophisticated brain scans, such as electromagnetic resonance, which can locate the site of a lesion while the patient is still alive.

The contributors to *Classic Cases in Neuropsychology* each describe the history of an important case or cases ranging from Broca to modern times, connecting them with recent research and theorizing. Neuropsychologists have shown that even higher cognitive processes are modular: that is, particular parts of the brain are involved in each. For example, some patients are unable to read, but their capacity to recognize objects other than words and their ability to write are only marginally impaired. Again, some exhibit a curious error of reading — when asked

to read words aloud they provide the wrong word, but one of the same semantic category as that presented: they may say 'chair' when shown 'table'. These mistakes therefore arise not at a perceptual level but after the word's meaning has been extracted. Again, there is the inability to recognize faces (prosopagnosia) while still being able to recognize certain aspects of them such as the direction of gaze.

Although neuropsychology has revealed *where* certain tasks are carried out in the brain, it has not yet thrown much light on *how* they are conducted, which is much the more interesting question. One notable exception is cases showing that the ability to store material in long-term memory can be abolished through brain damage while leaving short-term memory intact (for instance, repeating a string of telephone numbers), combined with a case in which short-term memory was eliminated but the person's long-term memory was preserved. This discredited the existing theory that material was passed from short-term to long-term memory, though, as so often, even this conclusion is not completely firm: it might be that the lesion made it impossible to retrieve material from short-term memory while the capacity to store such material briefly and pass it on to long-term memory was intact. Indeed, despite the fascination of these cases, there are few whose explanation has not remained controversial. Some believe that prosopagnosia can be an isolated effect, others that other forms of visual recognition are affected. It is difficult if not impossible to find lesions limited to an area that subserves a single function. Moreover, as one contributor warns, "Patients always present with a large variety of symptoms, of which only a minor portion are usually described in case studies".

Most of the contributors limit themselves to brain lesions in humans, but such work should be interpreted in the context of other approaches, including work on animals, which has the advantage that it is possible to make more precise lesions and also to stimulate or record directly from the brain. On the other hand, animals have the unfortunate disadvantage that they can neither speak nor write. There is, moreover, little mention of brain scans, which can be used to measure blood flow: if blood flow increases in a particular area while the subject is performing a certain task, then that area is probably implicated in the task.

Finally, because the book is organized around specific cases, some of the important findings of neuropsychology are ignored. For example, different lesions can destroy the ability to deal with different and specific categories: a lesion to part of the motor area selectively damages the capacity to recognize tools, suggesting that such recognition is at least partially

based on a knowledge of how to use them.

Classic Cases in Neuropsychology is an interesting and formidably scholarly book, but it is not easy reading and will appeal mainly to specialists. Some chapters correct errors that have crept into the literature, particularly from early German work that has never been translated. Given their aims, the editors have done a good job, covering aphasia, alexia, agnosia and so on, although a few of their contributors appear to be suffering from agraphia themselves. □

Stuart Sutherland is in the Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

Niche fillers

G. E. Fogg

Algae: An Introduction to Phycology. By C van den Hoek, D. C. Mann and H. J. Jahns. Cambridge University Press: 1996. Pp. 623. £70, \$110 (hbk); £24.95, \$39.95 (pbk).

THE rag-bag of organisms grouped as algae ranges in size over eight orders of magnitude and includes both prokaryotes and eukaryotes with little in common except oxygenic photosynthesis. Nevertheless, phycology has persisted as a distinct science and calls for textbooks surveying the whole of its catholic field.

Half a century ago, F. E. Fritsch managed to put nearly everything known about algae into a scholarly treatise, using the theme of parallel evolution in the different classes to give the work coherence. Since then, electron microscopy and molecular biology have provided insights leading not just to the overhaul of classification but also to the realization that the algae display a major evolutionary radiation at the cellular level. The phycologist's instinct that these organisms should be studied as a single group has been spectacularly vindicated.

Algae is an updated version of a highly successful German text. It presents traditional life histories and morphology, together with salient features of ecology and physiology, but concentrates on modern concepts of phylogenesis. It is excellently illustrated, mostly using clear line drawings of fine structure rather than electron micrographs.

Phycologists will keep Fritsch's volumes on their shelves, but this new book will displace others alongside and take over the wear and tear of everyday use by both student and research worker. □

C. E. Fogg is at the Marine Science Laboratories, Menai Bridge, Gwynedd LL59 5EY, UK.

New in paperback

Reflections of Eden: My Life with the Orangutans of Borneo by Biruté M. F. Galdikas. Indigo, £7.99. "Unlike the books by her fellow trinitates [Jane Goodall and Dian Fossey], *Reflections of Eden* is highly egocentric, sometimes frustratingly so", W. C. McGrew, *Nature* **374**, 832 (1995).

Charles Darwin: Voyaging — Volume 1 of a Biography by Janet Browne. "Her qualifications as a trained biologist, historian of science and skilled editor of the correspondence put her in an ideal position.... A wonderful read", P. R. Sloan, *Nature* **374**, 829 (1995).

Feral Children and Clever Animals: Reflection on Human Nature by Douglas Keith Candland. Oxford University Press, £12.99. \$16.95. "An eye-opening history of psychology", Andrew Whiten, *Nature* **366**, 375 (1993).

Identification of a major co-receptor for primary isolates of HIV-1

HongKui Deng^{*}, Rong Liu[†], Wilfried Ellmeier^{*}, Sunny Choe[†], Derya Unutmaz^{*‡}, Michael Burkhardt^{*}, Paola Di Marzio[†], Shoshana Marmon^{*}, Richard E. Sutton[§], C. Mark Hill^{*}, Craig B. Davis^{*}, Stephen C. Peiper^{||}, Thomas J. Schall[¶], Dan R. Littman^{*‡} & Nathaniel R. Landau[†]

^{*} Skirball Institute for BioMolecular Medicine, and [‡] Howard Hughes Medical Institute, New York University Medical Center, New York, New York 10016, USA

[†] Aaron Diamond AIDS Research Center, Rockefeller University, New York, New York 10016, USA

[§] Department of Biochemistry, Stanford University Medical Center, Stanford, California 94305, USA

^{||} James Graham Brown Cancer Center, University of Louisville School of Medicine, Louisville, Kentucky 40207, USA

[¶] Department of Immunology, DNAX Research Institute, Palo Alto, California 94304, USA

Entry of HIV-1 into target cells requires cell-surface CD4 and additional host cell cofactors. A cofactor required for infection with virus adapted for growth in transformed T-cell lines was recently identified and named fusin. However, fusin does not promote entry of macrophage-tropic viruses, which are believed to be the key pathogenic strains *in vivo*. The principal cofactor for entry mediated by the envelope glycoproteins of primary macrophage-tropic strains of HIV-1 is CC-CKR-5, a receptor for the β -chemokines RANTES, MIP-1 α and MIP-1 β .

THE human immunodeficiency viruses infect CD4⁺ macrophages and T helper cells¹. Although HIV-1 entry requires the cell-surface expression of CD4, to which the viral envelope glycoprotein (Env) binds, several studies have suggested that it is not sufficient for fusion of the viral envelope to the cellular plasma membrane. Early studies showed that, although human cells expressing a transfected CD4 gene were permissive for virus entry, murine cells expressing human CD4 were not²⁻⁴. These findings led to the suggestion that a species-specific cell-surface cofactor is required, in addition to CD4, for HIV-1 entry. Subsequent studies showed that strains of HIV-1 that had been adapted for growth in transformed T-cell lines (T-tropic strains) could infect primary T cells, but not primary monocytes or macrophages. In contrast, many primary viral strains were found to infect monocytes, macrophages and primary T cells, but not transformed T-cell lines⁵⁻⁸. This difference in tropism was found to be a consequence of specific sequence differences in the gp120 subunit of Env^{9,10}, suggesting that multiple cell-type-specific cofactors may be required for entry, in addition to CD4.

The nature of the cofactors required for HIV entry proved elusive until the recent identification of fusin¹¹, a member of the seven-transmembrane G-protein-coupled receptor family. Fusin was shown to act as a co-receptor for T-tropic strains, but it did not support syncytium formation of CD4⁺ cells with cells expressing Env of macrophage-tropic viruses. These viruses more closely resemble those that predominate in infected individuals throughout the course of the disease, particularly in the asymptomatic phase. In addition, these strains appear to be responsible for HIV-1 transmission, both sexually and by the transfer of infected blood^{12,13}. Rare individuals who are resistant to sexual transmission of HIV-1 have T cells that are readily infected by T-tropic virus, but cannot be infected by macrophage-tropic virus, further supporting a role for macrophage-tropic virus in the transmission of HIV-1 (ref. 14).

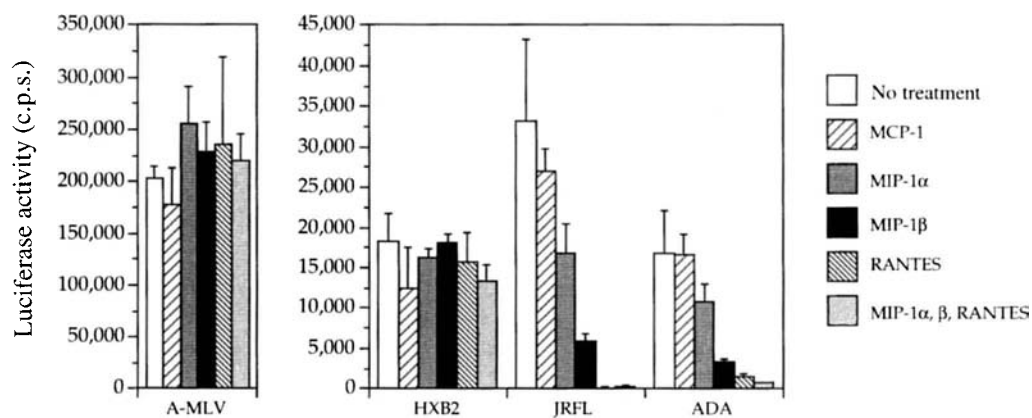
The inhibitors of HIV-1 replication present in supernatants of CD8⁺ T cells have recently been characterized¹⁵ as the β -chemokines RANTES (regulated-upon-activation, normal T expressed and secreted), macrophage inflammatory protein-1 α (MIP-1 α)

and MIP-1 β . Chemokines are chemotactic cytokines that are released by a wide variety of cells to attract macrophages, T cells, eosinophils, basophils and neutrophils to sites of inflammation (reviewed in refs 16,17). There are two classes of chemokines, C-X-C (α) and C-C (β), depending on whether the first two cysteines are separated by a single amino acid (C-X-C) or are adjacent (C-C). The α -chemokines, such as interleukin-8 (IL-8), neutrophil-activating protein-2 (NAP-2) and melanoma growth stimulatory activity protein (MGSA) are chemotactic primarily for neutrophils, whereas β -chemokines, such as RANTES, MIP-1 α , MIP-1 β , monocyte chemotactic protein-1 (MCP-1), MCP-2, MCP-3 and eotaxin are chemotactic for macrophages, T-cells, eosinophils and basophils. The chemokines bind specific cell-surface receptors belonging to the family of G-protein-coupled seven-transmembrane-domain proteins (reviewed in ref. 18). On binding their cognate ligands, chemokine receptors transduce an intracellular signal through the associated trimeric G protein, resulting in a rapid increase in intracellular calcium concentration. There are at least seven human chemokine receptors that bind or respond to β -chemokines with the following characteristic pattern: CC-CKR-1 (refs 19, 20) (MIP-1 α , MIP-1 β , MCP-3, RANTES), CC-CKR-2A and CC-CKR-2B (MCP-1, MCP-3), CC-CKR-3 (ref. 21) (eotaxin, RANTES, MCP-3), CC-CKR-4 (ref. 22) (MIP-1 α , RANTES, MCP-1), CC-CKR-5 (ref. 23) (MIP-1 α , RANTES, MIP-1 β), and the Duffy blood-group antigen²⁴ (RANTES, MCP-1).

Several lines of evidence implicate chemokine receptors as possible accessory factors in infection by primary strains of HIV-1. First, fusin is a member of the seven-transmembrane-domain family of chemokine receptors. It is most closely related to the IL-8 receptor, having a homology of 39% in the transmembrane domains²⁵. Presumably, fusin is a receptor for some currently unknown chemokine or neuropeptide. Second, the finding that the β -chemokines RANTES, MIP-1 α and MIP-1 β inhibit infection by primary or macrophage-tropic HIV-1, but not by T-tropic virus¹⁵, suggests that chemokine receptors are involved in HIV-1 replication and implicates the macrophage-tropic Env in this process. Third, CD4⁺ cells of individuals that have been multiply

FIG. 1 Chemokines block infection at the level of viral entry. PM1 cells were infected with luciferase reporter viruses pseudotyped by Envs of HIV-1 macrophage-tropic (ADA and JRFL) or T-cell-line adapted virus (HXB2) or with A-MLV Env in the presence or absence of individual β -chemokines or a mixture of MIP-1 α , MIP-1 β and RANTES. Luciferase activity was measured 4 days later, as described below. This experiment was repeated 4 times with similar results.

METHODS. NL4-3-Luc-R⁻E⁻²⁷ virus stocks pseudotyped by various Envs were generated by transfecting 293T cells with 10 μ g each of pNL4-3-Luc-R⁻E⁻ and pcDNA1/Amp-based expression vectors (Invitrogen) encoding JRFL, ADA²⁸, BaL²⁹, HXB2 or amphotropic MLV Env³⁰. Virus-containing supernatants were collected 48 h post-transfection and frozen at -80°C . Viruses were quantified by p24 ELISA assay. Cells (5×10^4) were seeded in 48-well dishes in DMEM containing 10% fetal bovine serum and infected with



luciferase reporter virus (50 ng p24) in a total volume of 400 μ l, with or without 30 min pretreatment with each of the chemokines listed (500 ng ml⁻¹, Peprotech). After 16 h, 0.5 ml medium was added to the wells. After 4 days of additional culture, 100- μ l lysates were prepared and luciferase activity in 20 μ l was assayed using commercially available reagents (Promega).

exposed to HIV-1 are highly resistant to infection *in vitro* by primary and macrophage-tropic strains of HIV-1 (ref. 14). Resistance to infection was correlated with an overproduction of chemokines. Taken together, these findings suggest a role for chemokines or chemokine receptors in the replication of primary but not T-cell-line adapted virus. These studies did not determine the step in the viral life cycle that was blocked by chemokines.

Here we report that β -chemokines inhibit HIV-1 replication by blocking entry of the virus into CD4⁺ cells. In light of this finding and those described above, we surmised that one or more of the β -chemokine receptors may serve as a required accessory factor for entry by macrophage-tropic HIV-1. We therefore tested the major members of the CC-CKR family for their ability to facilitate infection with macrophage-tropic HIV-1 strains and fusion with cells expressing Env proteins from these strains. Our results show that the product of the recently identified gene encoding CC-CKR-5 acts together with CD4 to allow entry of primary macrophage-tropic strains of HIV-1.

Chemokines block entry of primary HIV-1

To test whether β -chemokines block entry of macrophage-tropic HIV-1, we infected the T cell line PM1, which is highly susceptible to infection with both macrophage-tropic and T-tropic virus²⁶, with HIV-1-based luciferase reporter viruses²⁷. The luciferase reporter viruses infect cells in a single round but are not competent for further replication because of a frameshift mutation inserted into the *env* gene. Thus measurement of luciferase activity in cells infected with pseudotypes of this virus allows comparison of the relative efficiency of entry mediated by different Envs. In these studies we used HXB2 as a representative T-tropic Env, and three commonly studied macrophage-tropic Envs, JRFL⁶, ADA²⁸ and BaL²⁹. To control for possible post-entry or nonspecific effects of β -chemokines, we also prepared virus pseudotyped with amphotropic murine leukaemia virus (A-MLV) Env³⁰.

The β -chemokines inhibited infection of PM1 cells with virus pseudotyped by macrophage-tropic Env (JRFL and ADA). In contrast, the β -chemokines had no effect on infection with virus bearing T-tropic (HXB2) or A-MLV Envs (Fig. 1). Strongest blocking was observed with RANTES, followed in order of effectiveness by MIP-1 β and MIP-1 α . We found that MCP-1, MCP-3 and eotaxin had no inhibitory effect (Fig. 1, and data not shown). This same order was observed in inhibition of primary HIV-1 replication by β -chemokines¹⁵. Taken together, these findings strongly suggest that β -chemokine inhibition of viral replica-

tion is due to a block to entry of macrophage-tropic, but not T-tropic, HIV-1.

CC-CKR-5 is a potent co-receptor

Given that β -chemokines block HIV-1 entry, and CC-CKRs belong to the same gene family as fusin¹¹, the recently described accessory factor for the entry of T-tropic virus, we surmised that macrophage-tropic viruses use a β -chemokine receptor for fusion and entry into target cells. To test this hypothesis, we expressed the known β -chemokine receptors or fusin in several human and murine cell lines, and then tested their relative infectivity using HIV-luciferase pseudotyped with the different Envs.

Human embryonic kidney 293T cells transiently co-transfected with CD4 and the different chemokine receptor expression vectors were readily infected with virus pseudotyped by amphotropic or T-tropic Env, but not by virus lacking Env (Fig. 2a). Cells transiently transfected with expression vectors for CD4 plus CC-CKR-1, CC-CKR-2B, CC-CKR-3 or CC-CKR-4 were resistant to infection with virus pseudotyped with macrophage-tropic envelopes when compared with vector-transfected control cells (Fig. 2a). However, cells coexpressing CD4 and CC-CKR-5 displayed increases of three to four orders of magnitude in sensitivity to infection with viruses pseudotyped by ADA, BaL or JRFL Env (Fig. 2a). Almost identical findings were made with CC-CKR-5 cDNAs amplified from three different individuals (data not shown).

Infection of the 293T cells expressing both CD4 and CC-CKR-5 was completely blocked by the anti-CD4 monoclonal antibody Leu-3a (Fig. 2b). In addition, when pcCD4 was omitted from the transfection, CC-CKR-5 failed to support virus entry (Fig. 2c). Taken together, these findings indicate that CC-CKR-5 and CD4 must function cooperatively to mediate entry of macrophage-tropic virus.

Murine cells transfected with human CD4 are resistant to infection with all tested strains of HIV. To determine whether chemokine receptors could confer susceptibility to infection, the different receptor genes were stably introduced into murine 3T3.CD4 cells. Cells expressing CC-CKR-1, CC-CKR-2B, CC-CKR-3, CC-CKR-4, Duffy or fusin were all resistant to infection with HIV-luciferase pseudotyped with macrophage-tropic Envs, but were infected with virus bearing amphotropic Env (Fig. 3a, and data not shown). Expression of CC-CKR-5 allowed infection with the macrophage-tropic pseudotypes, but these cells were resistant to infection mediated by HXB2 Env (Fig. 3a). Only fusin-expressing 3T3.CD4 cells were permissive for infection with this

T-tropic virus (Fig. 3a). The β -chemokine receptors were expressed on the surface of the 3T3.CD4 cells, as assessed by mobilization of intracellular free Ca^{2+} in response to the appropriate chemokines (Fig. 3b, and data not shown). Cells expressing CC-CKR-5 responded to RANTES, MIP-1 α and MIP-1 β , which is consistent with known β -chemokine reactivities²³. Infection of the 3T3.CD4 cells expressing CC-CKR-5 with macrophage-tropic virus was blocked by a mixture of the three chemokines that efficiently activate this receptor, as well as by anti-CD4 antibody (Fig. 3a). Infection of the fusin-expressing cells with T-tropic virus was also blocked by anti-CD4, but was completely refractory to treatment with chemokines. Thus these results suggest that only CC-CKR-5 mediates entry of macrophage-tropic Envs, that T-tropic envelope glycoproteins do not use this co-receptor for entry, and that β -chemokines block entry of the macrophage-tropic virus by specifically binding to this receptor.

Stable expression of CC-CKR-5, but not of the other β -chemokine receptors, in human HOS.CD4, HeLa.CD4 and U87MG.CD4 (ref. 31) cells also conferred susceptibility to infection with macrophage-tropic HIV-1 (Fig. 3c, d, and data not shown). As observed in the transient transfections, stable coexpression of both CC-CKR-5 and CD4 was required for viral entry into the HeLa cells (Fig. 3d). Infection of these cells with macrophage-tropic virus was reduced by 70–80% when they were treated with a mixture of chemokines (Fig. 3d). Unexpectedly, high levels of β -chemokines failed to inhibit infection of HOS.CD4 cells (data not shown). In general, inhibition with β -chemokines was consistently less efficient in the non-lymphoid cells expressing CD4 and CC-CKR-5 than in the PM1 cells.

CC-CKR-5 promotes Env-mediated fusion

Fusion of the HIV-1 envelope with the cellular plasma membrane can be simulated by co-cultivating cells that express Env with human cells that express CD4, thus resulting in formation of syncytia³. Murine cells expressing human CD4 fail to support this fusion. Expression of fusin renders murine cells fusogenic for cells expressing T-tropic, but not macrophage-tropic, Env¹¹. To test whether CC-CKR-5 would support fusion with cells expressing macrophage-tropic Env, we transfected 293T cells with different Env expression vectors and co-cultivated them overnight with cell lines stably expressing transfected CD4 and CC-CKR-5 genes. As shown in Fig. 4, 293 T cells expressing JRFL Env formed large syncytia with murine 3T3.CD4 cells expressing CC-CKR-5, but

not with cells expressing fusin. Conversely, 293T cells expressing HXB2 Env were found to fuse to cells expressing fusin, but not to cells expressing CC-CKR-5. Similar results were obtained with U87MG.CD4 cells transfected with either fusin or CC-CKR-5 (not shown). Thus macrophage-tropic Env-mediated fusion occurs in a manner that is highly specific for the entry cofactor.

Replication of macrophage-tropic virus

To test whether CC-CKR-5 expression allows for full replication and spread of macrophage-tropic virus, we infected HOS.CD4 cells expressing CC-CKR-5 and control cells (HOS.CD4-BABE, transduced with the puromycin-resistance vector alone) with the replication-competent reporter viruses HIV-HSA (HSA, heat-stable antigen)³² and HIV(BaL)-HSA. These viruses contain the gene for HSA (CD24) in place of *nef*, allowing for quantification of the infected cells by fluorescence-activated cell sorting (FACS) after staining with anti-HSA monoclonal antibody³². Both viruses are based on the T-cell-line adapted virus NL4-3, but the latter contains the BaL macrophage-tropic Env. Both replicated in PM1 cells (Fig. 5a), but HIV(BaL)-HSA fails to replicate in T-cell lines, such as CEMX174, and in HOS.CD4 (data not shown).

The HOS.CD4-BABE cells remained uninfected with both viruses six days after infection, but nearly all of the HOS.CD4-CKR5 cells were infected with HIV(BaL)-HSA (Fig. 5a, bottom panel). Sampling of the HIV(BaL)-HSA-infected cultures over a period of several days indicated that an increasing percentage of the cells became infected over time, confirming the ability of the virus to spread in the culture (Fig. 5b). HIV-HSA failed to replicate in the HOS.CD4-CKR5 cultures, consistent with the restriction of this T-tropic virus to using fusin, which is expressed at very low levels in these cells (data not shown). Expression of CC-CKR-5 in 3T3.CD4 cells also led to HIV(BaL)-HSA virus replication, but this was rather limited, presumably owing to inefficient viral gene expression in murine cells (data not shown).

Expression of CC-CKR-5

The initial description of the CC-CKR-5 gene suggested that its expression is limited to granulocyte precursors, and absent in peripheral blood mononuclear cells (PBMCs)²³. To be a major co-receptor *in vivo*, however, this molecule should be expressed in T cells and monocyte/macrophages, the predominant cell-types targeted by the virus. Northern blot analysis with CC-CKR-5 cDNA as probe does not readily distinguish between CC-CKR-5

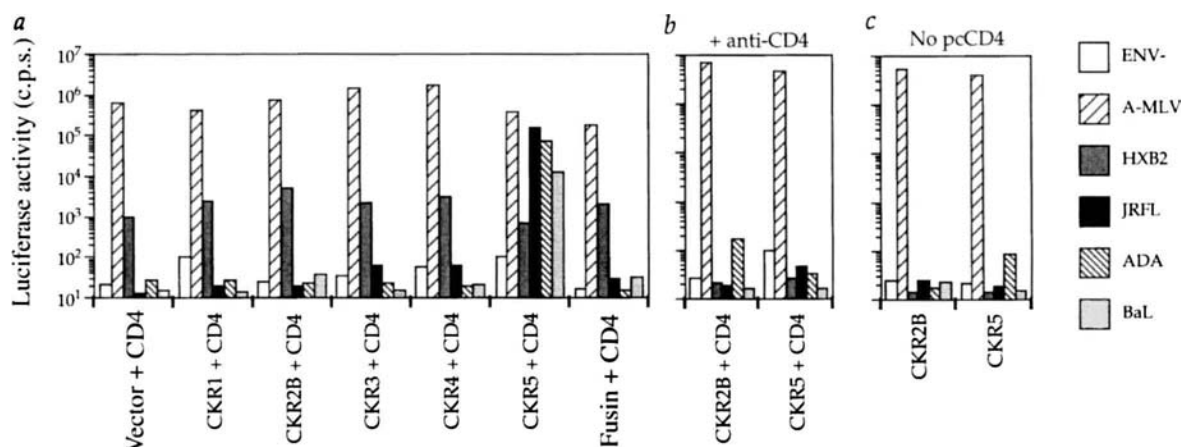


FIG. 2 CC-CKR-5 mediates entry of macrophage-tropic HIV-1. cDNAs encoding CC-CKR-1, CC-CKR-2B, CC-CKR-3, CC-CKR-4 and CC-CKR-5 were amplified from activated PBMC RNA using primers hybridizing to the respective 5' and 3' untranslated regions. Amplified products were cloned into pcDNA-I/Amp and pBABE-puro⁴⁰ expression vectors. Each of the cDNAs was sequenced and determined to correspond to that previously reported. a, Transfection of 293T cells with 5 μ g CD4 expression vector pcCD4⁴¹ and

15 μ g pcDNA-I expression vectors for each of the CC-CKR genes. The next day the cells were plated in 24-well dishes (2×10^4 per well), and one day later were infected with luciferase reporter viruses (20 ng p24) in a volume of 300 μ l. Luciferase activity was measured 4 days later, as described above. b, As a, with addition of anti-CD4 monoclonal antibody Leu3a (20 μ g ml⁻¹) 30 min before adding virus. c, As a, except that pcCD4 was omitted from the transfection and replaced by pcDNA-I control vector DNA.

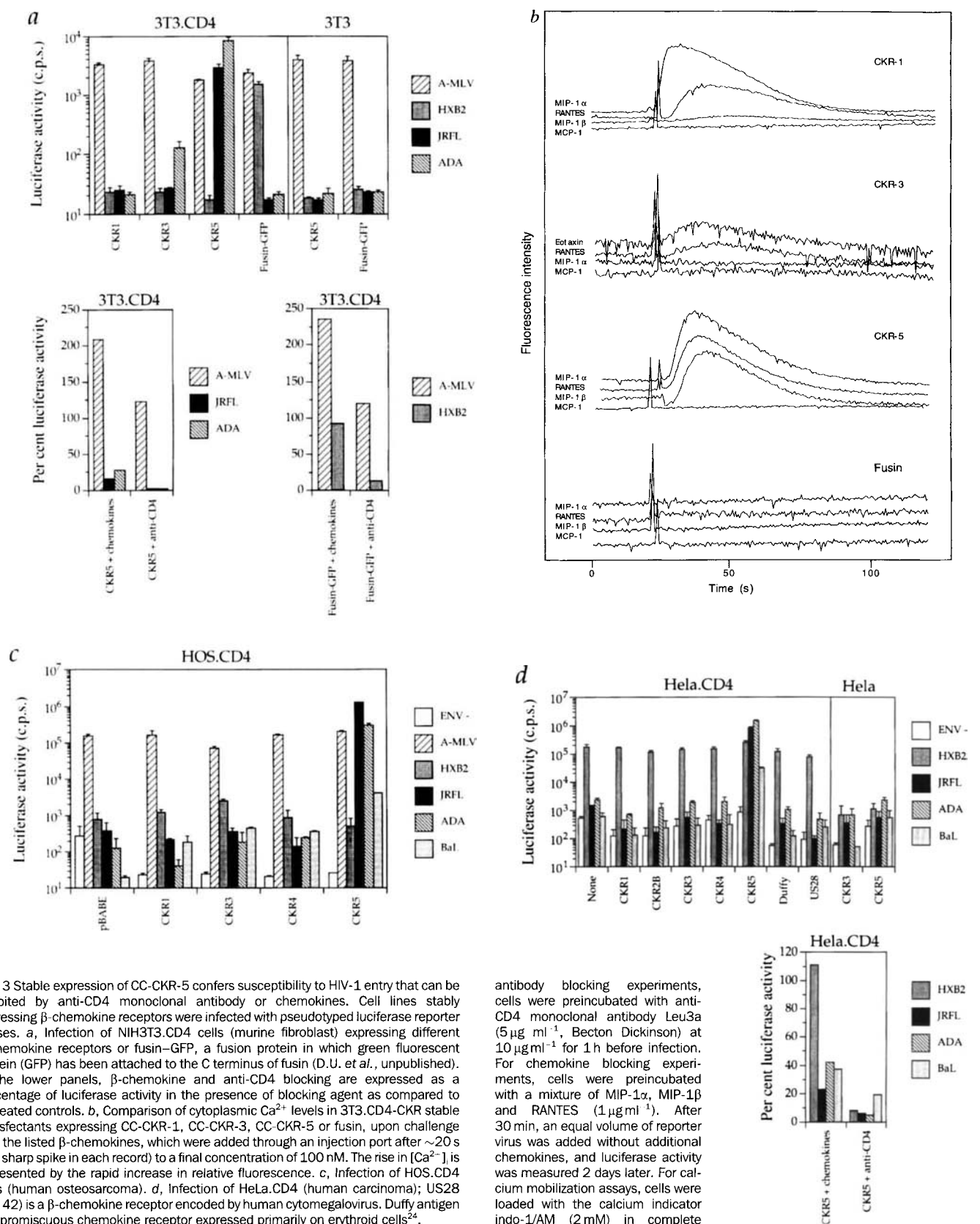


FIG. 3 Stable expression of CC-CCR-5 confers susceptibility to HIV-1 entry that can be inhibited by anti-CD4 monoclonal antibody or chemokines. Cell lines stably expressing β -chemokine receptors were infected with pseudotyped luciferase reporter viruses. **a**, Infection of NIH3T3.CD4 cells (murine fibroblast) expressing different β -chemokine receptors or fusin-GFP, a fusion protein in which green fluorescent protein (GFP) has been attached to the C terminus of fusin (D.U. *et al.*, unpublished). In the lower panels, β -chemokine and anti-CD4 blocking are expressed as a percentage of luciferase activity in the presence of blocking agent as compared to untreated controls. **b**, Comparison of cytoplasmic Ca^{2+} levels in 3T3.CD4-CCR stable transfectants expressing CC-CCR-1, CC-CCR-3, CC-CCR-5 or fusin, upon challenge with the listed β -chemokines, which were added through an injection port after ~ 20 s (the sharp spike in each record) to a final concentration of 100 nM. The rise in $[Ca^{2+}]_i$ is represented by the rapid increase in relative fluorescence. **c**, Infection of HOS.CD4 cells (human osteosarcoma). **d**, Infection of HeLa.CD4 (human carcinoma); US28 (ref. 42) is a β -chemokine receptor encoded by human cytomegalovirus. Duffy antigen is a promiscuous chemokine receptor expressed primarily on erythroid cells²⁴. **METHODS**. Cell lines stably expressing chemokine receptors or fusin-GFP were established as described^{43,44}. Briefly, cDNAs encoding the indicated receptors were subcloned into pBABE-puro⁴⁰. Amphotropic virus stocks were prepared by transfecting BING packaging cells⁴⁴ with the resulting plasmids or by a previous method^{43,44}, with the substitution of 293T cells for COS cells. Supernatants were collected 48 h later and used to infect NIH3T3 (3T3), 3T3.CD4, HOS, HOS.CD4, HeLa, and HeLa.CD4 cells. After another 48 h, cells were selected in medium containing $1 \mu\text{g ml}^{-1}$ puromycin. One week later, puromycin-resistant populations were tested for infectability by pseudotyped luciferase reporter virus (100 ng p24 per infection). For

antibody blocking experiments, cells were preincubated with anti-CD4 monoclonal antibody Leu3a ($5 \mu\text{g ml}^{-1}$, Becton Dickinson) at $10 \mu\text{g ml}^{-1}$ for 1 h before infection. For chemokine blocking experiments, cells were preincubated with a mixture of MIP-1 α , MIP-1 β and RANTES ($1 \mu\text{g ml}^{-1}$). After 30 min, an equal volume of reporter virus was added without additional chemokines, and luciferase activity was measured 2 days later. For calcium mobilization assays, cells were loaded with the calcium indicator indo-1/AM (2 mM) in complete growth medium at 20°C for 45 min. Cells were washed, resuspended in Na-HBSS (in mM: 2 $CaCl_2$, 145 NaCl, 5 KCl, 1 $MgCl_2$, 5 D-glucose, 20 HEPES, pH 7.3) containing 1% BSA, at 20°C for up to 2 h. Fluorescence measurements to determine $[Ca^{2+}]_i$ used a spectrofluorimeter (Photon Technologies) with $\sim 5 \times 10^6$ cells suspended in 2 ml Na-HBSS, at 37°C in a constantly stirred acrylic cuvette. The excitation wavelength was 350 nm (4 nm bandwidth) and dual simultaneous monitoring of emission was at 405 and 485 nm (10 nm bandwidth). The ratio of emission at 405/485 nm was measured at a rate of 2 Hz.

and the closely related CC-CCR-2 transcripts, so we performed reverse-transcriptase polymerase chain reaction (RT-PCR) on isolated subsets from PBMCs. CC-CCR-5 transcripts were detected in both the monocyte/macrophage and macrophage-depleted T-cell fractions (Fig. 6a). In addition, we found that PM1 and HUT78 cells both expressed the gene. Significantly more CC-CCR-5 transcript was detected in PM1 cells, consistent with the higher infectivity of these cells by macrophage-tropic and primary HIV-1 isolates²⁶.

Discussion

Here we have shown that CC-CCR-5 acts as a potent coreceptor that, together with CD4, allows entry of macrophage-tropic HIV-1 into cells. Both CD4 and CC-CCR-5 are required, just as CD4 and fusin are required for the entry of T-cell-line adapted virus¹¹. Co-receptor usage appears to be highly sequence specific. Other β -chemokine receptor family members tested showed no detectable co-receptor activity, including the closely related CC-CCR-2, which is 76% identical to CC-CCR-5 (ref. 23).

Several lines of evidence suggest that CC-CCR-5 serves as a major co-receptor for primary macrophage-tropic strains of HIV-1 *in vivo*. The three independently derived macrophage-tropic Envs used in our study were specific for CC-CCR-5. These Envs were derived from virus after limited growth in PBMCs and are therefore likely to maintain co-receptor usage similar to that of primary virus. We also found that a variety of human and murine cells transfected with human CD4 and CC-CCR-5 expression vectors became highly prone to infection by macrophage-tropic virus. Thus these proteins are likely to function as co-receptors in cells in which they are coexpressed *in vivo*. Finally, the importance of macrophage-tropic virus in HIV-1 transmission³³ and the demonstration of high levels of β -chemokines in T-cell cultures of individuals resistant to infection¹⁴ further imply a key role for CC-CCR-5 *in vivo*.

Although our data support a central role for CC-CCR-5 *in vivo*, we cannot rule out the possible significance of other potential co-receptors. CC-CCR-5 seemed to be expressed in T lymphocytes and macrophages, but it remains possible that these cells express a currently unidentified co-receptor. Furthermore, although the macrophage-tropic Envs that we tested used only CC-CCR-5, infected individuals could harbour viruses specific for other members of the β -chemokine receptor family. Resolution of these issues will await definitive demonstration that virus entry into macrophages and T cells can be prevented by treatment with CC-CCR-5-specific antibodies.

The identification of CC-CCR-5 as a macrophage-tropic virus co-receptor, together with the recent identification of fusin as a co-receptor for T-tropic viruses¹¹, suggests a resolution for the long-standing puzzle of Env-related differences in HIV-1 tropism (as shown in Fig. 6b). Thus adaptation of primary HIV-1 isolates for growth in transformed T-cell lines may be the result of selection for Env sequences that use fusin rather than CC-CCR-5, presumably owing to loss of expression of the latter in transformed cells. *In vivo*, the phenotypic switch from macrophage-tropic, non-syncytium-inducing (NSI) to T-tropic, syncytium-inducing (SI) viruses that occurs in many infected individuals before an increase in the severity of the disease^{34,35} would reflect a change in predominant co-receptor usage from CC-CCR-5 to fusin. Although the driving force for this change in receptor usage has yet to be determined, it could be the result of selection for viruses that can replicate in the presence of high levels of β -chemokine, or that can infect a wider variety of cell types. With the availability of cell lines expressing the co-receptors it will now be possible to evaluate the receptor usage of viruses sampled at different stages of HIV disease progression.

The V3 loop of gp120 clearly plays a central role in determining viral tropism. Alteration of only a few amino acids of the loop alters the ability of the virus to infect macrophages or transformed T-cell lines^{29,36,37}. Presumably, the V3 loop plays a critical role in determining which co-receptor is used. These findings lead us to

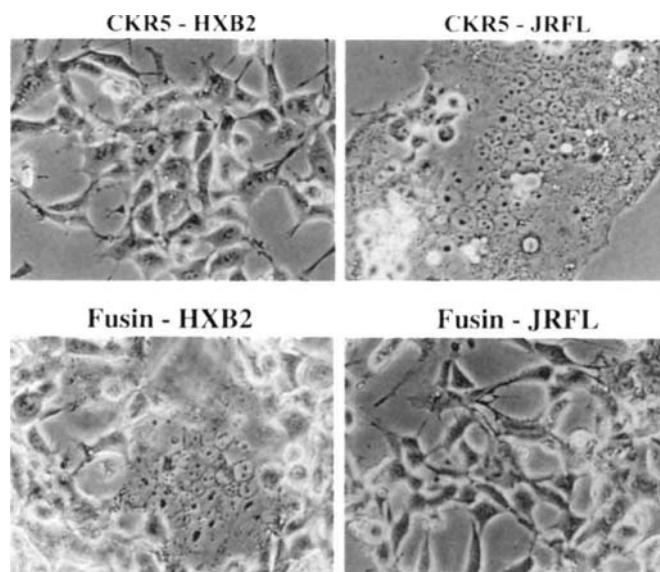
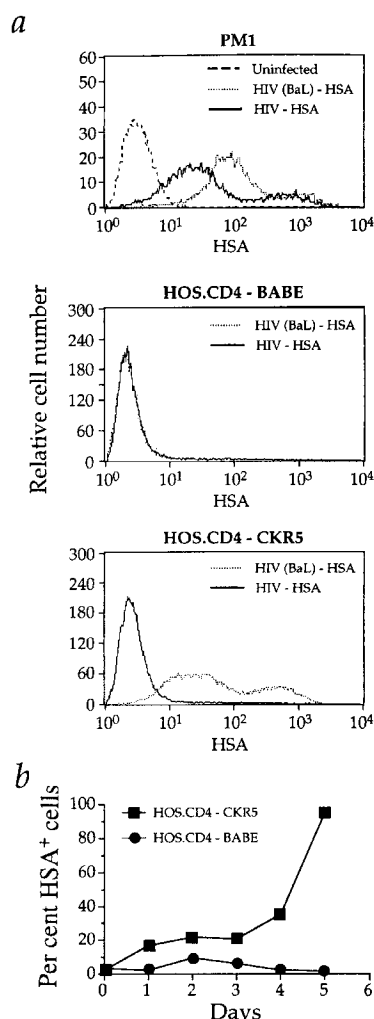


FIG. 4 CC-CCR-5 mediates Env-dependent fusion. We transfected 293T cells with equal amounts of pcDNA-I-based Env and Rev expression vectors. After 2 days the transfected cells (1.5×10^5) were seeded with 3T3.CD4-CKR5 or 3T3.CD4-fusin (3.0×10^5) cells. The next day the cells were stained with Giemsa stain and photographed.

FIG. 5 CC-CCR-5 supports

macrophage-tropic, but not T-cell-line adapted, virus replication in human cells. **a**, PM1, HOS-CD4-BABE and HOS-CD4-CKR-5 cells (5×10^5) were plated in 6-well dishes, and the next day infected with replication competent T-cell-line adapted HIV-HSA³² or macrophage-tropic HIV(BaL)-HSA reporter viruses (50 ng p24). HIV-HSA is based on the T-cell-line adapted virus pNL4-3, but contains, in place of *nef*, the gene encoding the small cell-surface protein, HSA (or CD24). HIV(BaL)-HSA virus is similar except that its *env* gene has been replaced by the *Sall*-*Bam*HI restriction fragment containing the macrophage-tropic Env of BaL. HIV(BaL)-HSA replicates in PM1 cells but not in CEM cells, while HIV-HSA replicates in both cell types (data not shown). Both viruses show a characteristic bimodal distribution of HSA-staining cells. This is likely to reflect whether the cells are in the early or late phase of the replication cycle. After 6 days the cells were stained with fluorescein isothiocyanate (FITC)- or phycoerythrin (PE)-conjugated anti-HSA monoclonal antibody (Pharmingen) and analysed in a Becton-Dickinson FACS-caliber. Both viruses failed to replicate in HOS-CD4-BABE cells, resulting in super-imposed curves. **b**, Time course of HIV(BaL)-HSA virus replicating in HOS-CD4-CKR5 cells. Cells were infected with HIV(BaL)-HSA and analysed by FACS on the indicated days.



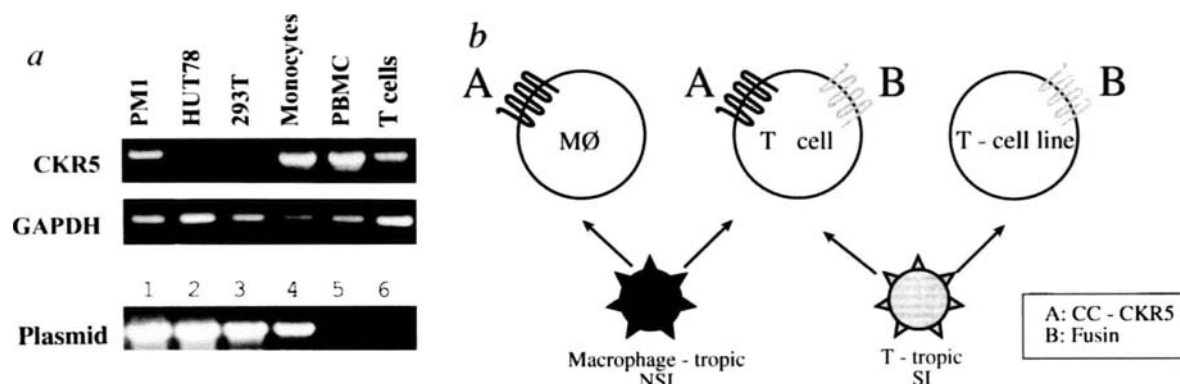


FIG. 6 CC-CCR-5 is expressed in T cells and monocyte/macrophages. *a*, Total RNA was prepared from the indicated cell-types using Triazol reagent (Gibco/BRL), treated with RNase-free DNase (Boehringer-Mannheim), and used in RT-PCR reactions. First-strand cDNA was primed with oligo-dT using Superscript reverse transcriptase (Gibco/BRL), and products were amplified with primers hybridizing to the 5' and 3' untranslated regions of CC-CCR-5 (upstream CTCGGATCCGTTGAACAAGATGGATTAT; downstream CTCGTCGACATGTGCACAACTCTGACTG) or to glyceraldehyde-3-phosphate dehydrogenase using a *Taq*/Pwo polymerase mixture (Boehringer Mannheim). To control for the presence of genomic DNA, control cDNA reactions in which reverse transcriptase was omitted were prepared in parallel. These

were uniformly negative (data not shown). To test the linearity of amplification, a 10-fold dilution series (lanes 1–5), starting at 1 pg of pcCCR5 plasmid DNA, was amplified under conditions identical to those above. In lane 6, no DNA was added. Monocytes were prepared by overnight adherence to plastic. T cells were prepared from the monocyte-depleted preparation by adherence to anti-CD2-coated beads (Dynal). *b*, Model for HIV-1 co-receptor usage. Macrophage-tropic viruses (dark shading), analogous to the NSI viruses that predominate early in infection, are proposed to use CC-CCR-5, which is expressed in macrophages and primary T cells. T-tropic viruses (light shading), analogous to SI viruses, use fusin, which is presumed to be expressed on primary T cells and transformed T-cell lines.

speculate that CD4 binding induces a conformational change in Env that exposes a co-receptor binding domain. This interaction site could either contain the V3 loop, or be conformationally influenced by it. In support of this model, increased exposure of the V3 loop upon CD4 binding to Env has been reported³⁸. A successful interaction with the co-receptor would trigger a conformational change in gp41, releasing its N-terminal hydrophobic peptide to initiate membrane fusion. Such a mechanism has a precedent in the low-pH-mediated activation of influenza haemagglutinin³⁹.

The identification of co-receptors required for HIV-1 entry raises several questions. The mechanism of chemokine blocking remains unclear: it may involve binding-site competition or

desensitization of the receptor through downregulation or conformational changes. The inefficient chemokine blocking that we observed with several cell lines suggests that competition for a binding site on the receptor may not be sufficient. In addition, it will be of interest to determine whether polymorphic differences in human *CC-CCR-5* alleles influence the susceptibility of individuals to HIV-1 transmission. It will also be important to determine whether signalling through the co-receptor is important for HIV-1 entry, for subsequent events in viral replication, or for elimination of T-helper cells. Transgenic mice that express human CD4 and the co-receptors may provide a means of addressing these questions. □

Received 22 May; accepted 31 May 1996.

- Sattentau, Q. J. & Weiss, R. A. *Cell* **52**, 631–633 (1988).
- Maddon, P. J. et al. *Cell* **47**, 333–348 (1986).
- Ashorn, P. A., Berger, E. A. & Moss, B. J. *Viol.* **64**, 2149–2156 (1990).
- Page, K. A., Landau, N. R. & Littman, D. R. *J. Virol.* **64**, 5270–5276 (1990).
- Gartner, S. et al. *Science* **233**, 215–219 (1986).
- Koyanagi, Y. et al. *Science* **236**, 819–822 (1987).
- Fisher, A. G. et al. *Nature* **334**, 444–447 (1988).
- Cheng-Mayer, C., Weiss, C., Seto, D. & Levy, J. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8575–8579 (1989).
- Liu, Z. Q., Wood, C., Levy, J. A. & Cheng-Mayer, C. *J. Virol.* **64**, 6148–6153 (1990).
- O'Brien, W. A. et al. *Nature* **348**, 69–73 (1990).
- Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. *Science* **272**, 872–877 (1996).
- Cornelissen, M. et al. *J. Virol.* **69**, 1810–1818 (1995).
- Veenstra, J. et al. *Clin. infect. Dis.* **21**, 556–560 (1995).
- Paxton, W. A. et al. *Nature Med.* **2**, 412–417 (1996).
- Cocchi, F. et al. *Science* **270**, 1811–1815 (1995).
- Schall, T. J. *Cytokine* **3**, 165–183 (1991).
- Murphy, P. M. A. *Rev. Immun.* **12**, 593–633 (1994).
- Horuk, R. *Trends pharmac. Sci.* **15**, 159–165 (1994).
- Ben-Baruch, A. et al. *J. biol. Chem.* **270**, 22123–22128 (1995).
- Neote, K., DiGregorio, D., Mak, J. Y., Horuk, R. & Schall, T. J. *Cell* **72**, 415–425 (1993).
- Combadere, C., Ahuja, S. K. & Murphy, P. M. *J. biol. Chem.* **270**, 16491–16494 (1995).
- Power, C. A. et al. *J. biol. Chem.* **270**, 19495–19500 (1995).
- Samson, M., Labbe, O., Mollereau, C., Vassart, G. & Parmentier, M. *Biochemistry* **35**, 3362–3367 (1996).
- Chaudhuri, A. et al. *J. biol. Chem.* **269**, 7835–7838 (1994).
- Jazin, E. E. et al. *Regul. Peptides* **47**, 247–258 (1993).
- Lusso, P. et al. *J. Virol.* **69**, 3712–3720 (1995).
- Connor, R. I., Chen, B. K., Choe, S. & Landau, N. R. *Virology* **206**, 936–944 (1995).
- Westervelt, P., Gendelman, H. E. & Ratner, L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3097–3101 (1991).

- (1991).
- Hwang, S. S., Boyle, T. J., Lyerly, H. K. & Cullen, B. R. *Science* **253**, 71–74 (1991).
- Landau, N. R., Page, K. A. & Littman, D. R. *J. Virol.* **65**, 162–169 (1991).
- Clapham, P. R., Blanc, D. & Weiss, R. A. *Circulation* **181**, 703–715 (1991).
- He, J. et al. *J. Virol.* **69**, 6705–6711 (1995).
- Zhu, T. et al. *Science* **261**, 1179–1181 (1993).
- Schuitmaker, H. et al. *J. Virol.* **66**, 1354–1360 (1992).
- Connor, R. I. & Ho, D. D. *J. Virol.* **68**, 4400–4408 (1994).
- De Jong, J. J., De Ronde, A., Keulen, W., Tersmette, M. & Goudsmit, J. *J. Virol.* **66**, 6777–6780 (1992).
- Fouchier, R. A. et al. *J. Virol.* **66**, 3183–3187 (1992).
- Sattentau, Q. J., Moore, J. P., Vignaux, F., Traincard, F. & Poignard, P. *J. Virol.* **67**, 7383–7393 (1993).
- Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. *Nature* **371**, 37–43 (1994).
- Morgenstern, J. P. & Land, H. *Nucleic Acids Res.* **18**, 3587–3596 (1990).
- Landau, N. R., Warton, M. & Littman, D. R. *Nature* **334**, 159–162 (1988).
- Gao, J. L. & Murphy, P. M. *J. biol. Chem.* **269**, 28539–28542 (1994).
- Landau, N. R. & Littman, D. R. *J. Virol.* **66**, 5110–5113 (1992).
- Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8392–8396 (1993).

ACKNOWLEDGEMENTS. We thank R. Connor, L. Ratner and I. Chen for HIV-1 Env constructs; P. Lusso for PM1 cells; S. Falkow for GFP constructs; S. Monard for FACS analysis; S. Martin and J. Li for technical assistance; and W. Chen and J. Weider for graphics. R.E.S. was supported by a Pfizer postdoctoral fellowship. H.-K.D. and P.D.M. were supported by fellowships from the Aaron Diamond Foundation. W.E. was supported by an Erwin Schrödinger postdoctoral fellowship. D.R.L. is an investigator of the Howard Hughes Medical Institute. This work was supported by grants from the National Institutes of Health to D.R.L. and N.R.L., and from the American Foundation for AIDS Research to N.R.L.

CORRESPONDENCE and requests for materials should be addressed to D.R.L. or N.R.L. (e-mail: littman@saturn.med.nyu.edu; nlandau@adarc.nyu.edu).

HIV-1 entry into CD4⁺ cells is mediated by the chemokine receptor CC-CKR-5

Tatjana Dragic*, Virginia Litwin†, Graham P. Allaway†, Scott R. Martin*, Yaoxing Huang*, Kirsten A. Nagashima†, Charmagne Cayanan*, Paul J. Maddon†, Richard A. Koup*, John P. Moore* & William A. Paxton*

* The Aaron Diamond AIDS Research Center, The Rockefeller University, New York, New York, 10016, USA

† Progenics Pharmaceuticals, Inc., Tarrytown, New York 10591, USA

The β -chemokines MIP-1 α , MIP-1 β and RANTES inhibit infection of CD4⁺ T cells by primary, non-syncytium-inducing (NSI) HIV-1 strains at the virus entry stage, and also block *env*-mediated cell-cell membrane fusion. CD4⁺ T cells from some HIV-1-exposed uninfected individuals cannot fuse with NSI HIV-1 strains and secrete high levels of β -chemokines. Expression of the β -chemokine receptor CC-CKR-5 in CD4⁺, non-permissive human and non-human cells renders them susceptible to infection by NSI strains, and allows *env*-mediated membrane fusion. CC-CKR-5 is a second receptor for NSI primary viruses.

THE replication of primary, non-syncytium-inducing (NSI) human immunodeficiency virus (HIV)-1 isolates in CD4⁺ T cells is inhibited by the β -chemokines macrophage inflammatory protein-1 α (MIP-1 α), MIP-1 β , and regulated-upon-activation, normal T expressed and secreted (RANTES)^{1,2}, but T-cell-line-adapted (TCLA) or syncytium-inducing (SI) primary strains are insensitive to these β -chemokines^{2,3}. CD4⁺ T cells from some HIV-1-exposed uninfected (EU) persons resist infection with NSI strains, but can be infected by TCLA and SI strains, and lymphocytes from some EU individuals secrete high concentrations of β -chemokines⁴. The β -chemokines are proteins of relative molecular mass 8,000 (M_r 8K). They are active on lymphocytes and monocytes by means of cell-surface receptors belonging to the family of G-protein-coupled seven-transmembrane-domain proteins⁴⁻⁸. One of these is the LESTR (also known as fusin) orphan receptor, the second receptor for TCLA HIV-1 strains⁹, which is not a receptor for known β -chemokines⁷⁻⁹.

β -Chemokines inhibit HIV-1 replication

To study how β -chemokines inhibit HIV-1 replication, we first used a virus entry assay based on single-cycle infection by an *env*-deficient virus, NL4/3 Δenv , which also carries the luciferase reporter gene, complemented by envelope glycoproteins expressed *in trans*^{10,11}. The use of PM1 cells, a variant of HUT-78 that supports replication of primary and TCLA HIV-1 strains, allowed comparison of *env* functions against a common cellular background^{2,12}. The β -chemokines MIP-1 α , MIP-1 β and RANTES are most active against HIV-1 in combination^{2,3}, and strongly inhibited infection of PM1 cells by viruses complemented with envelopes from the NSI strains ADA and BaL (Table 1a). Individually, RANTES and MIP-1 β were more strongly active than the other β -chemokines tested¹³⁻¹⁵ (Table 1a). MIP-1 α , MIP-1 β and RANTES in combination did not inhibit infection of PM1 cells by the TCLA strains NL4/3 and HxB2 (Table 1a). Thus phenotypic characteristics of the HIV-1 envelope glycoproteins influence their sensitivity to β -chemokines in a virus entry assay.

EU CD4⁺ T cells and NSI virus entry

The *env*-complementation assay was used to assess HIV-1 entry into CD4⁺ T cells from two individuals, EU2 and EU3, which are exceptionally resistant to infection by NSI strains in conventional HIV-1 infection assays³. The cells of neither individual supported

efficient entry of the NSI strain, JR-FL, but both allowed HxB2 entry; luciferase activity in cells from EU2 and EU3 after exposure to JR-FL was 300 and 200 c.p.m., respectively, compared with 5,440 and 29,560 c.p.m. from the same cells infected with HxB2. In contrast, JR-FL-infected CD4⁺ T cells from control individuals LW4 and LW5 produced luciferase counts of 814,670 and 77,880 c.p.m. The cells of EU2 and EU3 were therefore capable of efficiently replicating the HIV-1 genome once virus entry had been achieved. MIP-1 α , MIP-1 β and RANTES strongly inhibited JR-FL infection of the CD4⁺ T cells of LW4 and LW5, and weakly reduced HxB2 infection of both LW and EU cells (Table 1c).

To examine whether EU2 and EU3 had a genetic or acquired block to infection, we isolated CD4⁺ T cell clones. All 21 clones from EU2 and the one clone isolated from EU3 produced high levels of β -chemokines (especially RANTES), irrespective of their Th phenotype. They were resistant to infection by the SF162 NSI strain, compared to 22 readily infectable CD4⁺ clones from LW4 and LW5 (Table 2). The SI variant of SF162, R3H³, was significantly more replication competent than SF162 in the EU clones, but the two strains replicated comparably in the LW clones. However, some EU clones resisted infection by both SF162 and R3H. That all the EU clones were essentially uninfected by SF162 suggested that the mechanism of resistance had a genetic basis. One possibility was that constitutive overproduction of β -chemokines in the CD4⁺ T cells of EU2 and EU3 rendered them resistant to infection. Anti- β -chemokine antibodies partly abolished resistance to HIV-1 SF162 infection of CD4⁺ T cells from LW4 when they were co-cultured with cells from EU2 (and hence were exposed to β -chemokines secreted from the cells of EU2). However, CD4⁺ cells from EU2 remained resistant to SF162 infection in the presence of these antibodies (Fig. 1a), suggesting that the resistance mechanism may be more complex than an overproduction of endogenous β -chemokines.

Inhibition early in HIV-1 infection

We determined when β -chemokines inhibited HIV-1 replication by showing that complete inhibition of the infection of PM1 cells required the continuous presence of β -chemokines for up to 5 h after the addition of BaL *env*-complemented HIV-1 (Fig. 1b). Pretreatment of the cells with β -chemokines for 2 or 24 h before infection had no inhibitory effect if the cells were subsequently washed before virus addition. Furthermore, adding β -chemokines

2 h after the virus only minimally affected virus entry (Fig. 1b). We next used an assay based on the polymerase chain reaction (PCR) to detect HIV-1 early-DNA reverse transcripts in PM1 cells after 10 h of infection. We found that reverse transcription of ADA, but not of NL4/3, was strongly inhibited in the presence of MIP-1 β and

RANTES (Fig. 1c). Thus inhibition by β -chemokines requires their presence during at least one of the early stages of HIV-1 replication: virus attachment, fusion and early reverse transcription.

We discriminated between these sites of action by testing

TABLE 1 Inhibition of HIV-1 entry and membrane fusion

	Luciferase activity (%)				Fusion (%)	
	BaL	ADA	NL4/3	HxB2	HeLa-JR-FL	HeLa-BRU
(a) PM1 cells						
Control without virus	2	2	2	5		
Control with virus	100	100	100	100	No chemokines	100
+R/M α /M β (50/50/50)	2	3	92	117	+R/M α /M β (80/400/100)	95
+RANTES (100)	1	1	n.d.	n.d.	+RANTES (80)	100
+MIP-1 α (100)	54	54	n.d.	n.d.	+MIP-1 α (400)	100
+MIP-1 β (100)	1	6	n.d.	n.d.	+MIP-1 β (100)	93
+MCP-1 (100)	46	50	n.d.	n.d.	+MCP-1 (100)	98
+MCP-2 (100)	28	26	n.d.	n.d.	+MCP-2 (100)	93
+MCP-3 (100)	58	46	n.d.	n.d.	+MCP-3 (100)	99
(b) Macrophages						
Control without virus	1	1	n.a.p.	n.a.p.	No chemokines	d.n.f.
Control with virus	100	100	n.a.p.	n.a.p.	+R/M α /M β (80/400/100)	d.n.f.
+R/M α /M β (50/50/50)	84	n.d.	n.a.p.	n.a.p.	+RANTES (80)	d.n.f.
+RANTES (100)	103	68	n.a.p.	n.a.p.	+MIP-1 α (400)	d.n.f.
+MIP-1 α (100)	93	90	n.a.p.	n.a.p.	+MIP-1 β (100)	d.n.f.
+MIP-1 β (100)	90	82	n.a.p.	n.a.p.	+MCP-1 (100)	d.n.f.
+MCP-1 (100)	107	90	n.a.p.	n.a.p.	+MCP-2 (100)	d.n.f.
+MCP-2 (100)	84	76	n.a.p.	n.a.p.	+MCP-3 (100)	d.n.f.
+MCP-3 (100)	88	96	n.a.p.	n.a.p.		
(c) CD4⁺ T cells						
JR-FL		HxB2			LW5	
LW4					No chemokines	100
Control without virus	1	1			+R/M α /M β (106/533/133)	100
Control with virus	100	100			+RANTES (106)	95
+R/M α /M β (200/200/200)	14	68			+MIP-1 α (533)	100
					+MIP-1 β (133)	92
LW5					+OKT4a (3 μ g/ml)	0
Control without virus	1	1			EU2	
Control with virus	100	100			No chemokines	100
+R/M α /M β (200/200/200)	15	73			+R/M α /M β (320/1600/400)	100
					+RANTES (320)	100
EU2					+MIP-1 α (1600)	100
Control without virus	n.a.p.	14			+MIP-1 β (400)	100
Control with virus	n.a.p.	100				
+R/M α /M β (200/200/200)	n.a.p.	51				
EU3						
Control without virus	n.a.p.	3				
Control with virus	n.a.p.	100				
+R/M α /M β (200/200/200)	n.a.p.	n.d.				

Effect of β -chemokines on HIV-1 entry and *env*-mediated membrane fusion. Virus entry assay (left). PM1 cells were cultured as described¹². Macrophages were isolated from peripheral blood by adherence to plastic in 96-well plates. The monolayer was washed extensively over 2 weeks to remove non-adherent cells and allow monocyte differentiation. Ficoll/hypaque-isolated peripheral blood mononuclear cells (PBMCs) from laboratory workers (LW) or EU individuals were stimulated with PHA for 72 h before depletion of CD8⁺ lymphocytes with anti-CD8 immunomagnetic beads (Dyna, Great Neck, NY). CD4⁺ lymphocytes were maintained in culture medium containing interleukin-2 (IL-2) (100 U ml⁻¹; Hoffmann LaRoche, Nutley, NY), as described³. Target cells (10⁵ to 2 $\times$ 10⁵) were infected with supernatants (10–50 ng of HIV-1 p24) from 293 cells cotransfected with an NL4/3 Δ *env*-luciferase vector and a HIV-1 *env*-expressing vector^{10,11}. We then added β -chemokines (R & D Systems, Minneapolis) to the target cells simultaneously with virus, at the final concentrations (ng ml⁻¹) indicated in parentheses in the first column. The β -chemokine concentration range was selected based on previous studies^{2,3}. After 2 h, the cells were washed twice with PBS, resuspended in β -chemokine-containing media, and maintained for 48–96 h. Luciferase activity in cell lysates was measured as described previously^{10,11}. The values indicated represent luciferase activity (c.p.m. per ng p24 per mg protein), expressed relative to that in virus-control cultures lacking β -chemokines (100%), and are the means of duplicate or sextuplicate determinations. Membrane fusion assay (right). CD4⁺ target cells (mitogen-activated CD4⁺ lymphocytes, PM1 cells or macrophages) were labelled with octadecyl rhodamine (Molecular Probes, Eugene, OR), and HeLa-JR-FL cells and HeLa-BRU cells (or control HeLa cells, not shown) were labelled with octadecyl fluorescein (Molecular Probes), overnight at 37 °C. Equal numbers of labelled target cells and *env*-expressing cells were mixed in 96-well plates and β -chemokines (or CD4 mAb OKT4a) were added at the final concentrations (ng ml⁻¹) indicated in parentheses in the first column. Fluorescence emission values were determined 4 h after cell mixing¹⁶. If cell fusion occurs, the dyes are closely associated in the conjoined membrane such that excitation of fluorescein at 450 nm results in RET and emission by rhodamine at 590 nm. Percentage fusion is defined as equal to 100 $\times$ [(exp RET – min RET)/(max RET – min RET)], where max RET is the percentage RET obtained when HeLa-Env and CD4⁺ cells are mixed, exp RET is the percentage RET obtained when HeLa-Env and CD4⁺ cells are mixed in the presence of fusion-inhibitory compounds, and min RET is the percentage RET obtained when HeLa cells (lacking HIV-1 envelope glycoproteins) and CD4⁺ cells are mixed. The percentage RET value has been defined¹⁶, and each is the mean of triplicate determinations. These values were, for HeLa-JR-FL and HeLa-BRU cells respectively: PM1 cells, 11.5%, 10.5%; LW5 CD4⁺ cells, 6.0%, 10.5%; EU CD4⁺ cells, 0.1%, 4.1%; macrophages, 4.3%, 1.2%. Abbreviations: n.d., not done (inhibition by combinations of β -chemokines was so weak that individual β -chemokines were not tested); n.a.p., NL4/3 and HxB2 replicated so poorly in macrophages, and JR-FL replicated so poorly in EU2 and EU3 CD4⁺ T cells, that no assessment of inhibition by β -chemokines was possible; R/M α /M β , RANTES + MIP – 1 α + MIP – 1 β ; d.n.f., EU2 CD4⁺ cells did not fuse with HeLa-JR-FL cells, and macrophages did not fuse with HeLa-BRU cells.

whether β -chemokines inhibited binding of JR-FL or BRU gp120 to soluble CD4, or of tetrameric CD4-IgG2 binding to HeLa cells expressing oligomeric JR-FL envelope glycoproteins¹⁶. No inhibition by any of the β -chemokines was found in either assay, whereas the CD4 monoclonal antibody OKT4a was strongly inhibitory (not shown). Thus β -chemokines inhibit a step after CD4 binding, when conformational changes in the envelope glycoproteins lead to fusion of viral and cellular membranes¹⁷. Cell-cell membrane fusion is also induced by the interaction between gp120 and CD4, and can be monitored directly by resonance energy transfer (RET) between fluorescent dyes incorporated in cell membranes¹⁶. OKT4a completely inhibits membrane fusion of PM1 cells with HeLa cells expressing the envelope glycoproteins of either JR-FL (HeLa-JR-FL) or BRU (HeLa-BRU), confirming the specificity of the RET assay¹⁶. RANTES and MIP-1 β (and, to a lesser extent, MIP-1 α) strongly inhibited membrane fusion of HeLa-JR-FL cells with PM1 cells, whereas fusion between PM1 cells and HeLa-BRU cells was insensitive to these β -chemokines (Table 1a).

Similar results were obtained with primary CD4⁺ T cells from LW5 (Table 1c), although higher concentrations of β -chemokines were required to inhibit membrane fusion in the primary cells than in PM1 cells. In marked contrast to the cells of LW5, the CD4⁺ T cells of EU2 did not fuse with HeLa-JR-FL cells, although they clearly fused with HeLa-BRU cells in a β -chemokine-resistant manner (Table 1c). The RET assay demonstrates that β -chemokines interfere with *env*-mediated membrane fusion, and establishes that envelope glycoproteins from a primary, NSI strain

cannot fuse with CD4⁺ T cells from an EU individual, giving an indication of how these cells resist HIV-1 infection.

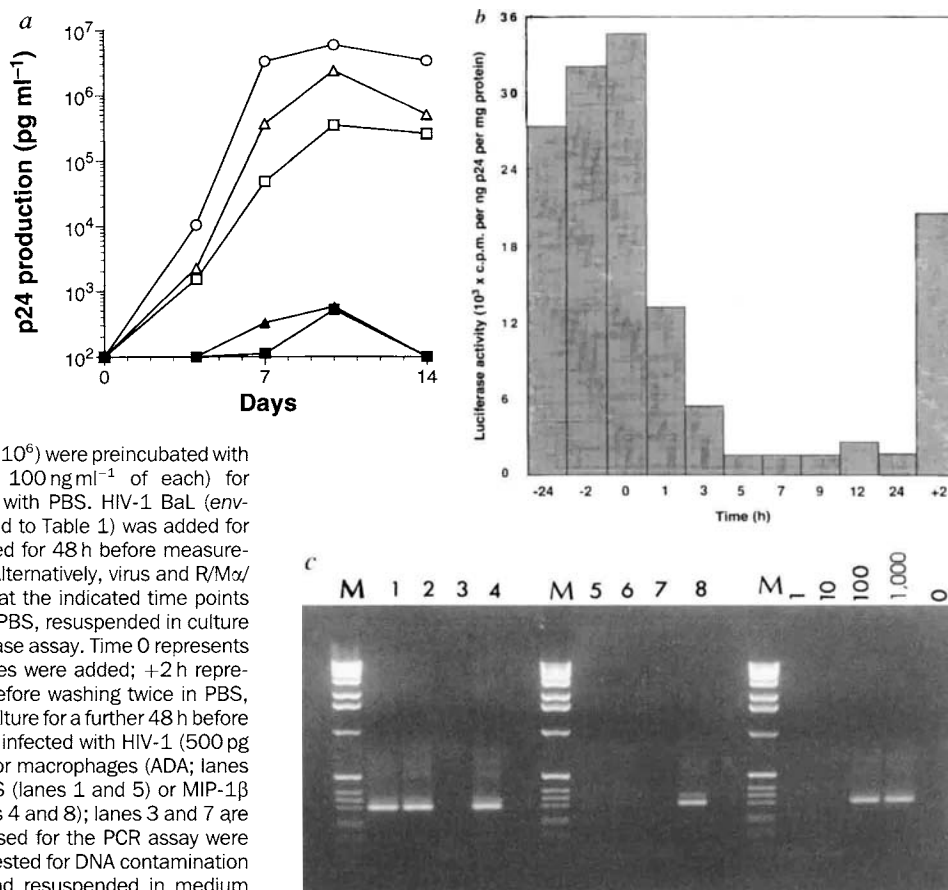
Cell-specific variables

Specific characteristics of envelope glycoproteins from different strains influence the susceptibility of HIV-1 to β -chemokines. There are, however, cell-type-specific variables to consider. For example, fusion of HeLa-JR-FL with human macrophages (normal donor) was only weakly inhibited by MIP-1 α , MIP-1 β , RANTES, MCP-2 and MCP-3, and not at all by MCP-1 (Table 1b). Similarly, infection of macrophages by BaL or ADA in the *env*-complementation assay was only minimally sensitive to very high concentrations (500 ng ml⁻¹) of MIP-1 α , MIP-1 β and RANTES (Table 1b, and data not shown). Hence, infection by the same virus (BaL or ADA) is either inhibited (T cells) or not inhibited (macrophages) by β -chemokines (Table 1); given the isogenic nature of the NL4/3 Δenv reporter virus^{10,11}, the target cell must be an important variable. Mitogen-activated primary CD4⁺ T cells are intermediate between macrophages and PM1 cells in their β -chemokine sensitivity, and there is also inter-donor variability. Thus BaL *env*-mediated entry was at 18, 36 and 101% of control levels in peripheral blood lymphocytes from three individuals in the presence of 50 ng ml⁻¹ each of MIP-1 α , MIP-1 β and RANTES.

CC-CKR-5 is a second receptor

The simplest explanation of our results is that the binding of certain β -chemokines to their receptor(s) prevents, directly or

FIG. 1 Specificity, time course and stage of β -chemokine inhibition of HIV-1 replication. a, Mitogen-activated CD4⁺ lymphocytes from LW4 (open circles), and equal numbers of CD4⁺ lymphocytes from LW4 and EU2 in the absence (open squares) or presence (open triangles) of polyclonal neutralizing antibodies to RANTES, MIP-1 α and MIP-1 β , or CD4⁺ lymphocytes from EU2 in the absence (filled squares) or presence (filled triangles) of these neutralizing antibodies, were inoculated with 600 TCID₅₀ of HIV-1 SF162. Virus replication was monitored by p24 antigen production³. Antibodies to RANTES, MIP-1 α and MIP-1 β (R&D Systems) were added at 200, 50 and 100 μ g ml⁻¹ respectively, and were replenished on day 7 of the culture. b, PM1 cells (1×10^6) were preincubated with RANTES + MIP-1 α + MIP-1 β (R/M α /M β ; 100 ng ml⁻¹ of each) for 24 h (-24) or 2 h (-2), then washed twice with PBS. HIV-1 BaL (*env*-complemented virus, 50 ng of p24; see legend to Table 1) was added for 2 h, then the cells were washed and incubated for 48 h before measurement of luciferase activity in cell lysates^{10,11}. Alternatively, virus and R/M α /M β were added simultaneously to cells, and at the indicated time points (1 h, 3 h, etc.) the cells were washed twice in PBS, resuspended in culture medium, and incubated for 48 h before luciferase assay. Time 0 represents the positive control, to which no β -chemokines were added; +2 h represents the mixture of virus with cells for 2 h before washing twice in PBS, addition of R/M α /M β and continuation of the culture for a further 48 h before luciferase assay. c, PM1 cells (1×10^6) were infected with HIV-1 (500 pg p24) grown in CEM cells (NL4/3; lanes 1-4) or macrophages (ADA; lanes 5-8), in the presence of 500 ng ml⁻¹ RANTES (lanes 1 and 5) or MIP-1 β (lanes 2 and 6), or with no β -chemokine (lanes 4 and 8); lanes 3 and 7 are negative controls (no virus). All viral stocks used for the PCR assay were treated with DNase for 30 min at 37 °C, and tested for DNA contamination before use. After 2 h, cells were washed and resuspended in medium containing the same β -chemokines for a further 8 h. DNA was then extracted from infected cells using a DNA/RNA isolation kit (US Biochemicals). First-round nested PCR was performed with primers: U3+, 5'-CAAGGCTACTTCCTGATTGGCAGAACTACACACAGG-3'; preGag, 5'-AGC-AAGCCGAGTCTCGCTCGAGAG-3'; and the second round with primers: LTR-test, 5'-GGGACTTTCGCTGGGACTTTC-3'; LRC2, 5'-CCTGTTCCGGCGCC-CTGCTAGAGATTTCCAC-3'; in a Perkin Elmer 2400 cyclo with the following



otherwise, the fusion of HIV-1 with CD4⁺ T cells. HIV-1 requires a second receptor for entry into CD4⁺ cells¹⁸⁻²⁰. This function is supplied, for TCLA strains, by LESTR⁹, so we considered that a related β -chemokine receptor might act as a second receptor for primary, NSI strains on CD4⁺ T cells. Several receptors for MIP-1 α , MIP-1 β and RANTES have been identified^{6,7}, and β -chemokines exhibit considerable cross-reactivity in receptor usage⁴⁻⁸. However, we identified CC-CKR-1 and, especially, CC-CKR-5 as the most likely candidates, based on their tissue expression patterns and ligand-binding properties^{4,7,8,15,21}. CC-CKR-1, CC-CKR-5 and LESTR are each expressed at the mRNA level in PM1 cells (data not shown). We therefore amplified by PCR, cloned and expressed these and other β -chemokine receptors.

The expression of CC-CKR-5 in HeLa-CD4 (human), COS-CD4 (simian) and 3T3-CD4 (murine) cells rendered each of them readily infectable by ADA and BaL in the *env*-complementation assay of HIV-1 entry (Table 3). LESTR, CC-CKR-2a, CKR-3 and CKR-4 could not substitute for CC-CKR-5 in this assay, and CC-CKR-1 expression permitted only very limited entry of ADA and BaL into HeLa-CD4 cells, but not into COS-CD4 or 3T3-CD4 cells (Table 3). However using a PCR-based assay, we have consistently observed a low level of NL4/3 and ADA entry into each of these CC-CKR-1-expressing cell lines (data not shown). The expression of LESTR in COS-CD4 and 3T3-CD4 cells allowed HxB2 entry in the *env*-complementation assay, and HxB2 readily entered untransfected (or control plasmid-transfected) HeLa-CD4

TABLE 2 Production of β -chemokines and infectability of CD4⁺ T-cell clones

Clone	Th	p24 (pg ml ⁻¹)			Chemokines (pg ml ⁻¹)	
		SF162	R3H	RANTES	MIP-1 α	MIP-1 β
EU2.1	1	0	0	3,427	0	227
EU2.2	2	0	1,123	3,009	526	1,153
EU2.3	1	0	5,341	1,880	0	34
EU2.4	1	0	1,661	1,998	23	42
EU2.5	1	83	0	15,951	156	1,253
EU2.6	1	8,090	442	4,630	156	618
EU2.7	1	0	53,478	3,055	0	156
EU2.8	1	0	44,864	2,168	0	118
EU2.9	1	0	1,940	1,368	0	68
EU2.10	1	0	0	2,939	83	202
EU2.11	1	0	84	1,602	0	98
EU2.12	1	0	725	3,590	361	1,988
EU2.13	1	0	46,894	1,483	0	34
EU2.14	1	0	42,326	2,055	17	88
EU2.15	1	0	113	3,872	78	164
EU2.16	0	0	780	28,701	7,778	5,344
EU2.17	1	207	39,030	2,726	0	96
EU2.18	1	0	797	2,505	28	251
EU2.19	0	0	0	1,435	127	997
EU2.20	0	237	0	4,454	195	1,483
EU2.21	2	0	0	1,071	237	126
EU3.1	2	2,063	0	8,341	445	1,855
LW4.1	1	105,730	70,222	185	48	436
LW4.2	0	127,286	44,936	642	170	490
LW4.3	0	65,018	84,528	699	36	412
LW4.4	2	28,922	42,326	148	22	65
LW4.5	1	150,940	95,488	246	0	0
LW4.6	0	61,594	99,834	246	29	81
LW4.7	1	114,188	154,024	47	16	25
LW4.8	0	11,871	1,077	782	130	201
LW4.9	2	39,974	4,704	43	30	30
LW4.10	n.p.	22,794	39,030	53	36	51
LW4.11	2	3,439	1,202	<156	90	116
LW4.12	2	49,722	40,550	273	27	337
LW4.13	n.p.	555,548	363,020	169	0	0
LW5.1	2	66,160	93,432	227	22	25
LW5.2	n.p.	104,912	159,208	280	0	0
LW5.3	2	46,836	169,850	336	0	23
LW5.4	0	72,222	44,726	829	124	229
LW5.5	n.p.	129,638	284,338	127	0	0
LW5.6	1	163,028	120,736	401	0	0
LW5.7	2	114,844	109,892	0	32	62
LW5.8	n.p.	91,732	251,338	151	0	0
LW5.9	1	134,924	280,600	<156	124	14

CD4⁺ lymphocytes were purified (>96% purity) from peripheral blood, then maintained in culture as described in the legend to Table 1. After one week, lymphocytes were labelled with anti-CD8 antibody and negatively sorted on a FACS Vantage cell sorter (Becton Dickinson, San Jose, CA) to further deplete CD8⁺ T cells. The purified cells were cloned at 10, 2.5 and 0.5 cells per well in the presence of allogeneic feeder cells, an anti-CD3 antibody and IL-2, and were periodically restimulated. The expression of both CD3 and CD4 on the clonally expanded cells was confirmed by FACS analysis, and some cells were tested for clonality by determining V β T-cell receptor usage using sequence-specific primers. The concentrations of interferon- γ (IFN- γ), IL-4, RANTES, MIP-1 α and MIP-1 β were determined in the supernatants from 1×10^6 cloned cells using commercial ELISA kits (R&D Systems). Clones secreting IFN- γ but not IL-4 were defined as Th1, those secreting IL-4 but not IFN- γ were defined as Th2 clones, and those secreting both cytokines were defined as Th0 clones. Cells (2×10^5) from each clone were inoculated with 200 TCID₅₀ of the NSI HIV-1 strain SF162 or of its SI variant SF162R3H³. HIV-1 replication was assessed by p24 antigen production on day 14, using a commercial ELISA (Abbott Labs, Abbott Park, IL)³. Abbreviations: n.p., clones that produced neither IFN- γ nor IL-4, for which no Th phenotype could be assigned.

cells (Table 3). Entry of BAL and ADA into all three CC-CKR-5-expressing cell lines was almost completely inhibited by the combination of MIP-1 α , MIP-1 β and RANTES, whereas HxB2 entry into LESTR-expressing cells was insensitive to β -chemokines (Table 3, and data not shown). These results show that CC-CKR-5 functions as a second receptor for some primary, NSI HIV-1 strains in a manner broadly analogous to that established for LESTR and TCLA strains⁹.

CC-CKR-5 allows membrane fusion

We confirmed that CC-CKR-5 functions as a second receptor in assays of *env*-mediated membrane fusion. Transient expression of

CC-CKR-5 in COS-CD4 and HeLa-CD4 cells allowed both cell lines to fuse strongly with HeLa-JR-FL cells (Fig. 2*c,d*), whereas no fusion occurred when control plasmids were used (Fig. 2*a,b*). Expression of LESTR instead of CC-CKR-5 did not allow either COS-CD4 or HeLa-CD4 cells to fuse with HeLa-JR-FL cells, but did allow fusion between COS-CD4 cells and HeLa-BRU cells (data not shown), consistent with previous results⁹. A quantitative end-point for cell-cell fusion was also obtained, β -galactosidase is expressed in the fused cells and stains them blue in the presence of the X-gal substrate²². The numbers of blue syncytia per well of Fig. 2*a* were: *a*, 7; *b*, 0; *c*, 3,340; *d*, 5,100.

The β -chemokine receptors were also tested in the RET fusion

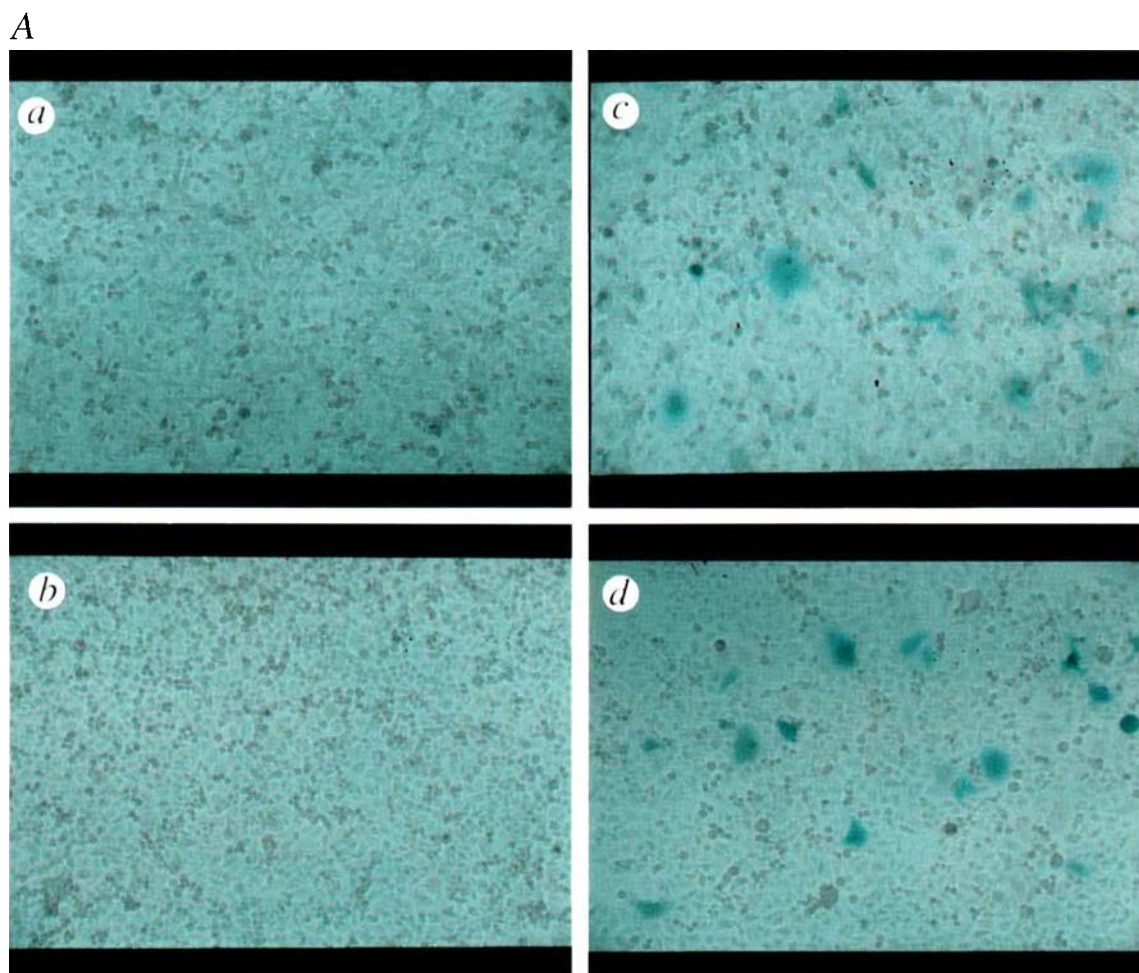
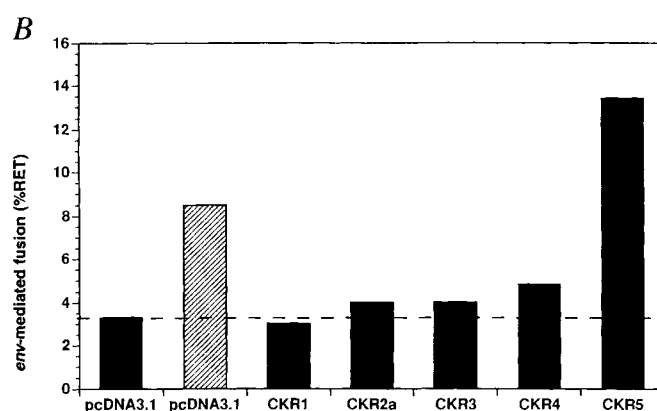


FIG. 2 HIV-1 *env*-mediated membrane fusion of cells transiently expressing C-C CKR-5. **A**, Membrane fusion mediated by C-C CKR-5 expression in HeLa and COS cells was demonstrated by transfecting control pcDNA3.1 into COS-CD4 (Z28T1) cells (*a*) or HeLa-CD4 (P42) cells (*b*), or transfecting pcDNA3.1-CKR5 into COS-CD4 cells (*c*) or HeLa-CD4 cells (*d*), then coculturing the transfected cells for 24 h with HeLa-JR-FL cells at a 1:1 ratio. The Z28T1 and P42 cells carry the inducible LTR-lacZ construct, and HeLa-JR-FL cells express HIV-1 tat. The β -gal staining was performed as described previously²². **B**, Membrane fusion mediated by β -chemokine receptors expressed in HeLa cells was demonstrated. Cells were transfected with control plasmid pcDNA3.1 or plasmid pcDNA3.1-CKR constructs using lipofectin (Gibco BRL). The pcDNA3.1 plasmid carries a T7-polymerase promoter, and transient expression of β -chemokine receptors was boosted by infecting cells with 1×10^7 plaque-forming units of vaccinia encoding the T7-polymerase (vFT7.3) 4 h post-lipofection⁹. Cells were then cultured overnight in R18-containing media, and tested for their ability to fuse with HeLa-JR-FL cells (black columns) or HeLa-BRU cells (hatched column) in the RET assay. Background fusion (pcDNA3.1 plasmid) is indicated by the broken line. The background level of fusion with control HeLa cells in this assay is typically between 3% and 4% RET, irrespective of



the transfected plasmid, so the percentage RET seen with C-C CKR-1, CKR-2a, CKR-3 and CKR-4 is not considered significant.

assay. Expression of CC-CKR-5, but not of CC-CKR-1, CKR-2a, CKR-3 or CKR-4, allowed fusion between HeLa-CD4 cells and HeLa-JR-FL cells to an extent greater than the constitutive level of fusion between HeLa-BRU cells and HeLa-CD4 cells (Fig. 2B). Fusion between HeLa-JR-FL cells and CC-CKR-5-expressing HeLa-CD4 cells was inhibited when RANTES, MIP-1 α and MIP-1 β (500 ng ml⁻¹ of each) were added at the same time as the fusion process was initiated by coculturing the cells. However, the extent of fusion inhibition by the β -chemokines depended on how long the HeLa-CD4 cells had been infected with the T7-polymerase-expressing vaccinia virus before they were co-cultured with the HeLa-JR-FL cells to initiate membrane fusion. Thus inhibition was 71, 46, 28 and 0% at 5, 7, 9 and 21 h, respectively, after infection by the vaccinia virus. This is probably because sustained overexpression of CC-CKR-5 overcomes the blocking effect of the CC-CKR-5 ligands.

Implications of second-receptor function

Taken together, our results establish that MIP-1 α , MIP-1 β and RANTES inhibit HIV-1 infection by interfering with the virus-cell fusion reaction subsequent to CD4 binding. Furthermore, we show that CC-CKR-5 can serve as a second receptor for entry of some primary NSI strains of HIV-1 into CD4⁺ T cells, and that CC-CKR-5 coexpression with CD4 allows membrane fusion mediated by the *env*-gene from NSI primary viruses. These findings are linked by the observation that the interaction of β -chemokines with CC-CKR-5 inhibits membrane fusion and HIV-1 entry. Overexpression of CC-CKR-5 in the membrane-fusion assay reduces the ability of β -chemokines to block fusion, which is suggestive of a competitive mechanism of inhibition. Further studies will be necessary to prove this, however, as other mechanisms, such as receptor downregulation, could also account for inhibition of CC-CKR-5-mediated fusion by β -chemokines.

We do not know if CC-CKR-5 is the only second receptor for NSI strains that can function on primary CD4⁺ T cells. Inter-donor variation in the β -chemokine sensitivity of HIV-1 infection suggests there may be some overlap in second receptor usage on activated CD4⁺ T cells by different NSI strains. Furthermore, the identity of the second receptor on macrophages is not yet known. Entry (and fusion) of NSI HIV-1 strains into macrophages was relatively insensitive to all β -chemokines tested, which contrasts with the observations from CD4⁺ T cells. Although another G-protein-coupled seven transmembrane domain protein may, perhaps, serve as a second receptor in macrophages, we do not rule out the possibility that CC-CKR-5 also fulfils this function on these cells. SI and TCLA strains do not enter or fuse with macrophages. Although it is not yet clear whether LESTR is expressed on macrophages, it was first cloned from macrophage/monocytes^{4,23}. There may be as-yet unexplained subtleties in second-receptor function and regulation in macrophages, and conceivably in other non-lymphoid cells. We note that the Duffy blood-group antigen on human erythrocytes is a promiscuous β -chemokine receptor^{24,25}. This might account for the observation that a heat- and protease-resistant factor present in human erythrocyte membranes can confer HIV-1 fusion competence to murine cells expressing human CD4, when heterokaryons of the two cell types are made^{26,27}. The Duffy antigen may not be a physiologically relevant second receptor, but might function *in extremis*.

Sequence changes in the *env* gene that occur during adaptation of HIV-1 to growth in cell lines, and conceivably those involved in the phenotypic switch^{28,29} from NSI to SI, might be the result of alterations in envelope glycoprotein configuration that allow the use of LESTR as well as, or instead of, CC-CKR-5 or a related receptor. It will be important to determine whether envelope glycoproteins bind directly to second receptors, and how this

TABLE 3 Infection of CD4-expressing cells by primary, NSI HIV-1 strains

		pcDNA3.1	LESTR	CKR-1	CKR-2a	CKR-3	CKR-4	CKR-5	R/M α /M β CKR-5
COS-CD4	ADA	798	456	600	816	516	534	153,000	3,210
	BaL	660	378	600	636	516	618	58,800	756
	HxB2	5,800	96,700	5,240	5,070	5,470	5,620	4,850	5,000
HeLa-CD4	ADA	678	558	4,500	912	558	600	310,000	6,336
	BaL	630	738	1,800	654	516	636	104,000	750
	HxB2	337,000	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	356,000
3T3-CD4	ADA	468	558	450	618	534	606	28,400	1,220
	BaL	606	738	660	738	534	558	11,700	756
	HxB2	456	24,800	618	672	732	606	618	606

The β -chemokine receptor genes C-C CKR-1, C-C CKR-2a, C-C CKR-3, C-C CKR-4 and C-C CKR-5 have no introns^{4-8,15,21}. They were isolated by PCR performed directly on a human genomic DNA pool derived from the PBMCs of 7 healthy donors. Oligonucleotides overlapping the ATG and the Stop codons, and containing *Bam*HI and *Xho*I restriction sites for directional cloning into the pcDNA3.1 expression vector (Invitrogen Inc.), were used. LESTR (also known as fusin or HUMSTR)^{4,9,23} was cloned by PCR performed directly on cDNA derived from PM1 cells, using sequences derived from the NIH database. All oligonucleotide sequences are listed below. PCR was performed in a Perkin Elmer 2400 cyclotherm with the following amplification cycles: 94 °C for 5 min, 30 cycles of {94 °C for 1 min, 60 °C for 1 min, 72 °C for 2 min}, 72 °C for 10 min. The human CD4-expressing cell lines HeLa-CD4 (P42), 3T3-CD4 (sc6) and COS-CD4 (Z28T1) (ref. 22) were transfected with the different pcDNA3.1-CKR constructs by the calcium-phosphate method, then infected 48 h later with different reporter viruses (200 ng HIV-1 p24 per 10⁶ cells) in the presence or absence of β -chemokines (400 ng ml⁻¹ each of RANTES, MIP-1 α and MIP-1 β). Luciferase activity in cell lysates was measured 48 h later^{10,11}. The β -chemokine blocking data are only shown for C-C CKR-5, as infection mediated by the other C-C CKR genes was too weak for inhibition to be quantifiable. We have not yet formally demonstrated that the clones C-C CKR genes are functional as β -chemokine receptors, so the inability of C-C CKR-1, CKR-2a, CKR-3, and CKR-4 to function as HIV-1 second receptors could be due to defective expression at the protein level. Listed below are the 5' and 3' primer pairs used in first (5-1 and 3-1) and second (5-2 and 3-2) round PCR amplification of the CKR genes directly from human genomic DNA, and of LESTR from PM1 cDNA. Only a single set of primers was used to amplify CKR-5. LESTR: L/5=AAG CTT GGA GAA CCA GCG GTT ACC ATG GAG GGG ATC; L/5-2=GTC TGA GTC TGA GTC AAG CTT GGA GAA CCA; L/3-1=CTC GAG CAT CTG TGT TAG CTG GAG TGA AAA CTT GAA GAC TC; L/3-2=GTC TGA GTC TGA GTC CTC GAG CAT CTG TGT. CKR-1: C1/5-1=AAG CTT CAG AGA GAA GCC GGG ATG GAA ACT CC; C1/5-2=GTC TGA GTC TGA GTC AAG CTT CAG AGA GAA; C1/3-1=CTC GAG CTG AGT CAG AAC CCA GCA GAG AGT TC; C1/3-2=GTC TGA GTC TGA GTC CTC GAG CTG AGT CAG. CKR-2a: C2/5-1=AAG CTT CAG TAC ATC CAC AAC ATG CTG TCC AC; C2/5-2=GTC TGA GTC TGA GTC AAG CTT CAG TAC ATC; C2/3-1=CTC GAG CCT CGT TTT ATA AAC CAG CCG AGA C; C2/3-2=GTC TGA GTC TGA GTC CTC GAG CCT CGT TTT. CKR-3: C3/5-1=AAG CTT CAG GGA GAA GTG AAA TGA CAA CC; C3/5-2=GTC TGA GTC TGA GTC AAG CTT CAG GGA GAA; C3/3-1=CTC GAG CAG ACC TAA AAC ACA ATA GAG AGT TCC; C3/3-2=GTC TGA GTC TGA GTC CTC GAG CAG ACC TAA. CKR-4: C4/5-1=AAG CTT CTG TAG AGT TAA AAA ATG AAC CCC ACG G; C4/5-2=GTC TGA GTC TGA GTC AAG CTT CTG TAG AGT; C4/3-1=CTC GAG CCA TTT CAT TTT TCT ACA GGA CAG CAT C; C4/3-2=GTC TGA GTC TGA GTC CTC GAG CCA TTT CAT. CKR-5: C5/5-12=GTC TGA GTC TGA GTC AAG CTT AAC AAG ATG GAT TAT CAA; C5/3-12=GTC TGA GTC TGA GTC CTC GAG TCC GTG TCA CAA GCC CAC.

might be achieved. The pattern of second receptor usage by HIV-1 strains of different genetic subtypes remains to be explored. However, the replication in peripheral blood lymphocytes of primary NSI strains from several genetic subtypes is inhibited by β -chemokines (A. Trkola and J.P.M., unpublished data), implying that viral phenotype is more important than genotype in determining β -chemokine sensitivity. Any binding site for the β -chemokine receptors on the envelope glycoproteins must, therefore, use highly conserved residues, or must have a strongly conserved

structural feature, such as charge, that is independent of major variations in primary sequence.

Finally, at least some EU individuals have CD4⁺ T cells that are relatively incompetent at fusing with NSI strains of HIV-1, probably because of a defect in second-receptor usage. Whether this is a result of autocrine ligation of CC-CKR-5 by overexpressed β -chemokines, or whether CC-CKR5 is dysfunctional for HIV-1 entry into these cells, has yet to be determined. □

Received 19 April; accepted 31 May 1996.

1. Levy, J. A., Mackewicz, C. E. & Barker, E. *Immun. Today* **17**, 217–224 (1996).
2. Coochi, F. et al. *Science* **270**, 1811–1815 (1995).
3. Paxton, W. A. et al. *Nature Med.* **2**, 412–417 (1996).
4. Neote, K., DiGregorio, D., Mak, J. Y., Horuk, R. & Schall, T. J. *Cell* **72**, 415–425 (1993).
5. Gao, J.-L. et al. *J. exp. Med.* **177**, 1421–1427 (1993).
6. Bacon, K. B., Premack, B. A., Gardner, P. & Schall, T. J. *Science* **269**, 1727–1729 (1995).
7. Raport, C. J. et al. *J. Leuk. Biol.* **59**, 18–23 (1996).
8. Wells, T. N. C. et al. *J. Leuk. Biol.* **59**, 53–60 (1996).
9. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. *Science* **272**, 872–877 (1996).
10. Chen, B. K., Saksele, K., Andino, R. & Baltimore, D. *J. Virol.* **68**, 654–660 (1994).
11. Connor, R. I., Chen, B. K., Choe, S. & Landau, N. R. *Virology* **206**, 935–944 (1995).
12. Lusso, P. et al. *J. Virol.* **69**, 3712–3720 (1995).
13. Charo, I. F. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 2752–2756 (1994).
14. Ben-Baruch, A. et al. *J. biol. Chem.* **270**, 22123–22128 (1995).
15. Combadiere, C. et al. *J. biol. Chem.* **270**, 29671–29675 (1995).
16. Litwin, V. et al. *J. Virol.* (in the press).
17. Moore, J. P., Jameson, B. A., Weiss, R. A. & Sattentau, Q. J. in *Viral Fusion Mechanisms* (ed. Bentz, J.) 233–289 (CRC Press, Boca Raton, FL, 1993).
18. Maddon, P. J. et al. *Cell* **47**, 333–348 (1986).
19. Ashorn, P. A., Berger, E. A. & Moss, B. *J. Virol.* **64**, 2149–2156 (1990).

20. Clapham, P. R., Blanc, D. & Weiss, R. A. *Virology* **181**, 703–715 (1991).
21. Samson, M., Labbe, O., Mollereau, C., Vassart, G. & Parmentier, M. *Biochemistry* **11**, 3362–3367 (1996).
22. Dragic, T., Charneau, P., Clavel, F. & Alizon, M. *J. Virol.* **66**, 4794–4802 (1992).
23. Loetscher, M. et al. *J. biol. Chem.* **269**, 232–237 (1994).
24. Chaudhuri, A. et al. *J. biol. Chem.* **269**, 7835–7838 (1994).
25. Neote, K., Mak, J. Y., Kolakowski, L. F. Jr & Schall, T. J. *Blood* **84**, 44–52 (1994).
26. Dragic, T., Picard, L. & Alizon, M. *J. Virol.* **69**, 1013–1018 (1995).
27. Puri, A., Morris, S. J., Jones, P., Ryan, M. & Blumenthal, R. *Virology* **219**, 262–267 (1996).
28. Connor, R. I. & Ho, D. D. *J. Virol.* **68**, 4400–4408 (1994).
29. Moore, J. P. & Ho, D. D. *AIDS* **9** (suppl. A), S117–S136 (1995).

ACKNOWLEDGEMENTS. We thank A. Trkola, D. Tse, S. Monard, R. Steinman, R. W. Doms and D. D. Ho for discussions; D. Tran and S. Choe for technical assistance; W. Chen and J. Guza for artwork; R. C. Gallo and P. Lusso for PM1 cells; C. Cheng-Mayer for virus stocks; B. Chen, R. I. Connor and N. Landau for luciferase reporter vectors; and L. Q. Zhang for primers. This work was funded by grants from the NIH, the Correlates of HIV Immune Protection contract, the Aaron Diamond Foundation, and by Progenics Pharmaceuticals, Inc. T.D. holds an Aaron Diamond fellowship and R.A.K. is an Elizabeth Glaser scientist of the Pediatric AIDS Foundation.

CORRESPONDENCE and requests for materials should be addressed to R.A.K. and J.P.M. (e-mail: koup@adarc.nyu.edu; moore@adarc.ngu.edu).

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

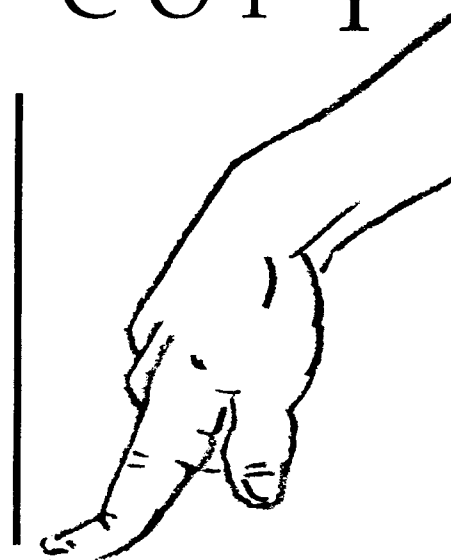
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

Smooth dark spiral arms in the flocculent galaxy NGC2841

David L. Block*, B. G. Elmegreen† & R. J. Wainscoat‡

* Department of Computational and Applied Mathematics, University of the Witwatersrand, Private Bag 3, WITS 2050, South Africa

† IBM Research Division, T. J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA

‡ Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

OPTICAL images of the arms of spiral galaxies invariably show massive blue stars forming in ridges of interstellar gas and dust¹. These are particularly striking in ‘grand-design’ galaxies, in which the stellar positions are influenced by spiral density waves’. By contrast, many galaxies have a ‘flocculent’ appearance, with no obvious evidence of spiral structure at visible wavelengths. Here we report infrared observations of the prototype flocculent galaxy NGC2841, which reveal a remarkable system of long, dark spiral arms. These arms arise from concentrations of dust; they are hidden at optical wavelengths by light scattered from the dust. The mechanism that has organized the gas and dust into these dark arms is at present unclear; the arms might be highly sheared dense clouds, or they might correspond to density waves in the interstellar medium driven by an elongated central bulge, which would not affect the stable stellar disk.

Optical images (Fig. 1a) of the galaxy NGC2841 reveal numerous patches of star formation and no grand-design spiral structure from a density wave in the underlying stellar disk^{1,2}. Such patchy structure occurs in over 60% of isolated non-barred galaxies³, giving them a flocculent (fleece-like) appearance. A characteristic of these galaxies is that the optical patches, by definition, span only a small range in azimuth. Flocculent structure probably results from a relatively low-mass disk compared to the halo⁴, and a higher than usual disk velocity dispersion^{5,6}; then stellar spiral waves are difficult to excite^{7,8} and most of the structure is from sheared features in the gas^{6,9,10}, such as bright star-forming regions or dark dusty clouds.

To search for density waves that may be hidden by dust in NGC2841, we imaged the galaxy in K' band¹¹ (2.1 μm) using a 1,024 $\times$ 1,024 pixel near-infrared camera and multiobject spectrograph (NICMOS) on the University of Hawaii 2.2-m telescope at Mauna Kea Observatory. The K' band highlights the red stars in the old stellar disk, which contains most of the disk mass, and it penetrates dust ten times better than visible light. A mosaic of 720-s exposure images smoothed to 0.7-arcsec resolution is shown in Fig. 1b and c. The distance to NGC2841 is assumed to be 9.5 Mpc, so 1 pixel = 0.7 arcsec = 32.2 pc.

No bright stellar waves are seen in the K' image. Usually such waves are obvious at K', where arm/interarm contrasts are a factor of 2 or more for the several dozen galaxies that have been imaged so far^{12–15}. The lack of bright K' spirals in NGC2841 implies that any waves that may be present are weaker than in grand-design galaxies by at least a factor of ten. This confirms our

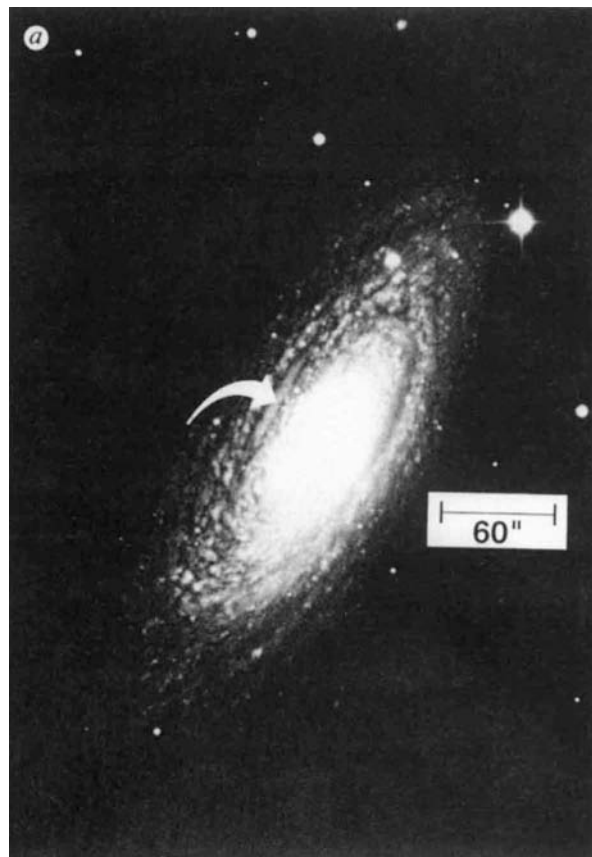
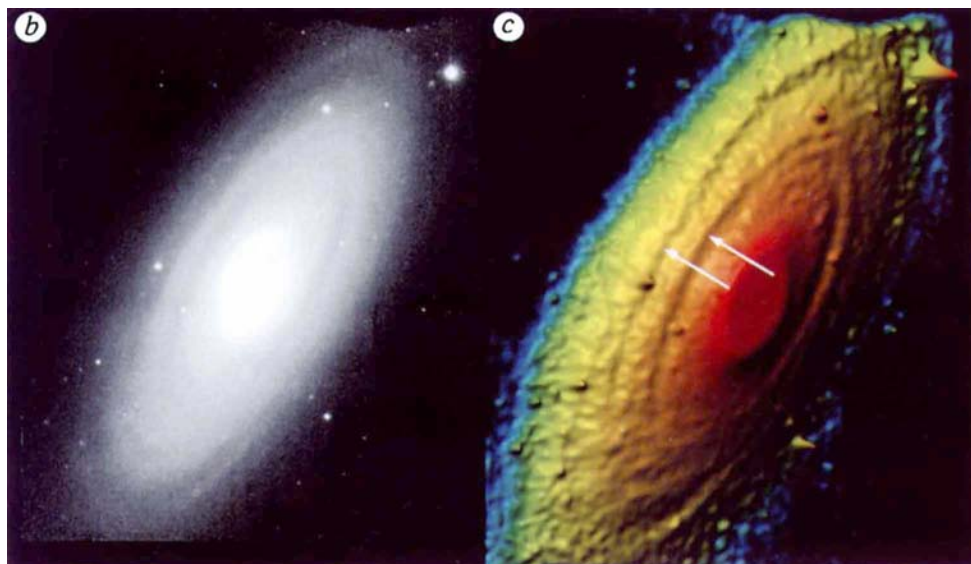


FIG. 1 a, Optical photograph of NGC2841 from ref. 2. Observer, M. Humason. North is up and east is to the left. The arrow points to a smooth amorphous strip of optical emission in the inner northeast dust lane (see text for details). b, c, K' image plotted as a grey-scale for log(intensity) (panel b) as a false-colour three-dimensional projection, also for log(intensity) (panel c). North is up and east is to the left. The bulge is clipped in the centre. The three-dimensional image is rotated slightly to highlight the shadows cast by an artificial light source. Two dark spirals are evident on the northeast minor axis (arrows), another on the southwest minor axis, and a fourth is in the south. Several other fainter dark spirals are present too. The bulge is misaligned 10° anticlockwise from the major axis of the outer disk.



impression from optical images that the disk is relatively stable against wave excitation.

A surprising feature of the K' image is the presence of long dark infrared spirals, two extending for 160° from the northeast minor axis to the north, and two others equally long in the west and south. Smaller dark spirals are present too. Most of these dark regions were not seen this clearly before (compare Fig. 1a), although the spiral on the southwest minor axis is faintly visible on the optical image in Wray's *Color Atlas of Galaxies*¹⁶.

How can a dust spiral be present in K' but not in a visible image where dust is usually more prominent? The answer is evident from Fig. 1a, which shows the dark spiral closest to the bulge in the northeast as a long, diffuse emission patch in visible light (see arrow in Fig. 1a), not as a dark patch. Such emission is very unusual for galaxies; it is not broken into stars or individual nebulae, nor is there a lack of dust (as it is dark on the K' image). Its location on the near side of the galaxy (considering the sense of rotation¹⁷ and the trailing spirals), just in front of the

bulge, makes it ideally positioned to scatter bulge light off the dense dust lane into our direction. This suggests that the most prominent dark spirals in the K' image are washed out by scattered light at shorter wavelengths.

The absorption and emission features on the near side of the galaxy are shown better in Fig. 2a. The diffuse visible (V) emission coming from the northeast inner dust spiral is evident at the position -27 arcsec from the nucleus. It is at slightly smaller projected radii than the K' extinction (the dip in the curve), which corresponds to either smaller physical radii or greater heights above the midplane compared with the centroid of the K'-band dark spiral (probably the latter for bulge-scattered light). The intensity of the excess diffuse emission is 23 counts, compared to 3,400 counts at the galaxy centre and 800 counts 2 kiloparsecs (kpc) from the centre on the minor axis; thus the dust lane forward-scatters about 1% of the bulge light it receives. We predict that the spectrum of the diffuse patch will contain stellar absorption lines from bulge stars.

To determine the opacity of the dust features, the K' intensity profile in Fig. 2a was fitted to a radiative transfer calculation (Fig. 2b) that incorporates absorption and phase-dependent scattering. The calculation integrates along lines of sight on the minor axis of a model exponential disk at the appropriate inclination of 68° . Model details are in the figure legend. The theoretical fit in Fig. 2b has K' extinctions through the inclined disk equal to 1.32 and 0.68 mag for the inner and outer dust spirals, and 0.84 and 0.30 mag for the ambient gas at the same radii without the spirals (1 mag is a factor of $10^{0.4} \approx 2.5$ in brightness or 0.92 optical depths). The mid-plane density in each dust spiral exceeds that in the surrounding regions by factors of four and eight, respectively. With these values, the V-band opacity (ten times higher than K') is high enough to obscure mid-plane stars and scatter bulge light off the top of the gas/dust layer; this makes the dark spiral on the nearside of the minor axis appear bright (as a smooth strip of emission) at V. In contrast, the K'-band opacity is only high enough to show the dust spirals—there is not enough column density for scattered light to obscure them at K'. This is why the long spiral arms in NGC2841 are dark at K' but indistinct or bright at visible wavelengths.

Observations of atomic^{17,18} and molecular^{19,20} gas at much poorer resolution than the K' images (arcminutes instead of arcseconds) confirm there is a reservoir of interstellar matter to make dark spirals out of a normal gas + dust mixture. The corresponding average V-band optical depth through the disk at the radii of the dark infrared spirals is ~ 2.5 mag, which is as high as for the famous dust screen in the 'Evil Eye' galaxy NGC4826 (ref. 21).

Figure 2c and d also indicates that the near side of the galaxy (northeast) has redder colours and sharper colour variations than the far side. The near-side dust lanes are very red, for example, and the patch of scattered light is relatively blue (red is upwards in these plots). The region of the bulge on both the major and minor axes is bluer than the near side, and the minor axis on the far side of the galaxy (positive distance offset in the figure) is the bluest. This blue gradient with distance is faintly evident in the Wray *Color Atlas*¹⁶ where the bulge and far-side minor axis look white and washed out compared to the near-side minor axis. The effect is much stronger in our data because the wavelength range covered by our V and K' bands is larger.

Such a colour gradient can be caused by scattered light from dust located high off the plane in the region of the bulge. In such a position, dust will selectively scatter blue bulge light at high angles into our line of sight and leave the direct paths redder; this makes the bulge and far-side bluer and more uniform than the near side. The fact that the blue excess occurs at a projected distance from the galaxy centre > 140 arcsec $= 6.4$ kpc implies that a substantial amount of dust has to be outside the deep potential well of the bulge—far above the disk. Radiation pressure or a nuclear wind may be required to get the dust this high.

Why are there dark spirals in NGC2841 if stellar density waves

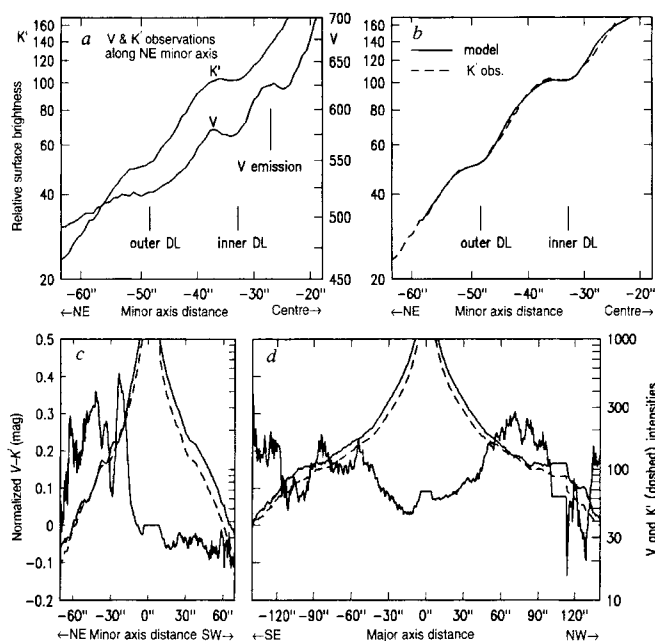


FIG. 2a, Intensity profiles in K' and V bands along a strip 29 arcsec wide on the northeast minor axis. The V-band data comes from a CCD (charge-coupled device) image with 0.7-arcsec resolution taken at the Wise Observatory by A. Heller. DL, dark lane b K'-band intensity profile (dashed line), and best-fit solid line using a radiative transfer model including absorption and scattering (solid line). The fit considers absorption and phase-dependent scattering by integrating along lines of sight on the minor axis of a model disk at an inclination of 68° . The disk has gaussian distributions of star light and dust with scale heights of 250 pc and 130 pc, and it has dust lanes with the same scale heights as the ambient gas and exponential distributions in the direction perpendicular to the arms towards larger and smaller radii. For scattering, we assume phase functions with various ratios of cross-sections in the forward, σ_{for} , average line-of-sight, σ_{los} , and backwards, σ_{back} , directions; the solution is determined by the proportion $\sigma_{\text{for}}/\sigma_{\text{los}}/\sigma_{\text{back}}$. We also assume that most of the ambient light that scatters into our view originates parallel to the plane. Then the essential quantity used in the calculation is the ratio $S = (\sigma_{\text{los}}/\sigma_s)/(1 + \sigma_a/\sigma_s - \sigma_{\text{for}}/\sigma_s)$ where σ_a and σ_s are the total absorption and scattering coefficients. The near-infrared albedo at K' is at least 0.5 or equivalently $\sigma_a/\sigma_s \approx 1$ (ref. 30 and references therein), which gives $S \approx 0.2$. c, Minor-axis V and K' profiles (centrally peaked curves) and V - K' relative magnitude difference. The central portion of the V image is saturated, so V - K' increases artificially toward the centre; this part is clipped, and the value there is arbitrarily set equal to zero. d, Major-axis V and K' profiles (centrally peaked curves) and V - K' relative magnitude difference, clipped in the centre, and normalized to the same zero point as the minor-axis profile. The colour gradient from the near side of the galaxy, on the northeast minor axis, to the centre on the major axis, to the far side on the southwest minor axis, is evident from a comparison of c and d.

are either weak or non-existent? Long dark spirals like these are unprecedented even for flocculent galaxies. Two possible reasons come to mind. One is related to the extremely large rate of shear in this galaxy, from differential rotation. At the position of the spirals, the rotation speed¹⁷ is a constant $V = 270 \text{ km s}^{-1}$, which is abnormally high for spiral galaxies like ours, and the deprojected radius is $R \approx 5 \text{ kpc}$, which is small compared to the galactocentric radius of the Sun (8.5 kpc) in the Milky Way. The corresponding rate of shear, $Rd(V/R)/dR$, is an enormous $54 \text{ km s}^{-1} \text{ kpc}^{-1}$. This is two times higher than the shear rate in the star-forming parts of the nearby Andromeda galaxy²³, its neighbour M33 (ref. 23), and in the solar neighbourhood of our Galaxy. Thus all of the gas features in NGC2841 should be much more spiral-like at the same evolutionary stage than they are in any of these nearby galaxies. We therefore propose that at least some of the dark spirals are simply sheared dark clouds. For a sheared region, the time elapsed from an initially circular state is $R/(V \tan i) = 1.1 \times 10^8 \text{ yr}$ for the observed pitch angle of $i = 9^\circ$ (measured from deprojected images). The mass of each dark spiral is $\sim 10^5 M_\odot$, from the density found in the radiative transfer solution. The dark spirals in NGC2841 are therefore similar in mass to the four largest dust clouds in the Andromeda galaxy²⁴. There are many smaller dark features in NGC2841 as well, which are presumably from smaller sheared clouds.

Another possibility is that some of the dark spirals are weak waves driven in the dusty interstellar medium by non-axisymmetric forces from an elongated bulge. Figure 1b indicates that the bulge in NGC2841 is not aligned along the major axis of the disk—the position angle differs by $\sim 10^\circ$. This means that the bulge is prolate if the disk is circular. Usually prolate bulges appear in barred²⁵ or grand-design galaxies such as Andromeda^{26,27} and NGC4845^{28,29}, where they can excite stellar waves²⁶, but here the stellar disk is apparently too stable for such excitation and the only response is in the dust and gas.

The distinction between sheared clouds and bulge-driven waves

should be evident from velocity measurements of the dark spiral gas. Systematic radial motions outwards or inwards on the minor axis would favour the wave interpretation. $\square$

Received 27 November 1995; accepted 30 April 1996.

- Elmegreen, D. M. & Elmegreen, B. G. *Astrophys. J. Suppl. Ser.* **54**, 127–150 (1984).
- Sandage, A. *Hubble Atlas of Galaxies* (Carnegie Instn, Washington DC, 1961).
- Elmegreen, D. M. & Elmegreen, B. G. *Mon. Not. R. astr. Soc.* **201**, 1021–1034 (1982).
- Elmegreen, D. M. & Elmegreen, B. G. *Astrophys. J.* **364**, 412–414 (1990).
- Bertin, G. in *Dynamics of Galaxies and Molecular Cloud Distributions* (eds Combes, F. & Casoli, F.) 93–104 (Kluwer Academic, Dordrecht, 1991).
- Elmegreen, B. G. & Thomasson, M. *Astr. Astrophys.* **272**, 37–58 (1993).
- Mark, J. W. K. *Astrophys. J.* **205**, 363–378 (1976).
- Toomre, A. in *The Structure and Evolution of Normal Galaxies* (eds Fall, S. M. & Lynden-Bell, D.) 111–136 (Cambridge Univ. Press, 1981).
- Seiden, P. E. & Gerola, H. *Astrophys. J.* **233**, 56–66 (1979).
- Elmegreen, B. G. & Efremov, Yu. *Astrophys. J.* (in the press).
- Wainscoat, R. J. & Cowie, L. L. *Astr. J.* **103**, 332–337 (1992).
- Block, D. L. & Wainscoat, R. J. *Nature* **353**, 48–50 (1991).
- Block, D. L. *et al.* *Astrophys. J.* **288**, 365–382 (1994).
- Rix, H. W. & Rieke, M. J. *Astrophys. J.* **418**, 123–134 (1993).
- Rix, H. W. & Zaritsky, D. in *Infrared Astronomy with Arrays: The Next Generation* (ed. McLean, I. S.) 151–152 (Kluwer Academic, Dordrecht, 1994).
- Wray, J. D. *The Color Atlas of Galaxies* (Cambridge Univ. Press, 1988).
- Bosma, A. *Astr. J.* **86**, 1825–1846 (1981).
- Begeman, K. G. thesis, Univ. Groningen (1987).
- Young, J. S. & Scoville, N. *Astrophys. J.* **260**, L41–L44 (1982).
- Braine, J. & Combes, F. *Astr. Astrophys.* **264**, 433–443 (1992).
- Block, D. L., Witt, A. N. & Grosbøl, P. in *The Opacity of Spiral Disks* (eds Davies, J. I. & Burstein, D.) 227–242 (Kluwer Academic, Dordrecht, 1995).
- Brinks, E. & Burton, W. B. *Astr. Astrophys.* **141**, 195–214 (1984).
- Deul, E. R. & van der Hulst, J. M. *Astr. Astrophys. Suppl. Ser.* **67**, 509–539 (1987).
- Hodge, P. W. *Astr. J.* **85**, 376–386 (1980).
- Wozniak, H., Friedli, D., Martinet, L., Martin, P. & Bratschi, P. *Astr. Astrophys. Suppl. Ser.* **111**, 115–152 (1995).
- Stark, A. A. *Astrophys. J.* **213**, 368–373 (1977).
- Lindblad, B. *Stockholm Obs. Ann.* **19**(2), 1–12 (1956).
- Bertola, F., Rubin, V. C. & Zeilinger, W. W. *Astrophys. J.* **345**, L29–L32 (1989).
- Gerhard, O. E., Vietri, M. & Kent, S. M. *Astrophys. J.* **345**, L33–L36 (1989).
- Witt, A. N., Lindell, R. S., Block, D. L. & Evans, Rh. *Astrophys. J.* **427**, 227–231 (1994).

ACKNOWLEDGEMENTS. Figure 1a is reproduced by permission of the Carnegie Institution of Washington and the California Institute of Technology. Figure 1b was made using the IBM Image Access Executive. Figure 1c was made using Data Explorer with the help of C. Pickover. The V image used for Fig. 2 was kindly supplied by A. Heller and N. Brosch. We thank S. Courteau, D. Elmegreen, and S. Federman for discussions, and F. Combes and K. G. Begeman for supplying gas column densities in tabular form.

Giant oxygen isotope shift in the magnetoresistive perovskite $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3+y}$

Guo-meng Zhao*, K. Conder†, H. Keller* & K. A. Müller*

* Physik-Institut der Universität Zürich, CH-8057 Zürich, Switzerland

† Laboratorium für Festkörperphysik ETH Zürich, CH-8093, Switzerland

FERROMAGNETIC perovskites of the form $\text{La}_{1-x}\text{Me}_x\text{MnO}_{3-y}$ (where Me is Ca or Sr) have been known¹ since 1950, but there has been a recent resurgence of interest following the discovery of giant magnetoresistance in this class of compounds^{2,3}. The compounds contain both Mn^{3+} and Mn^{4+} ions; as the electronic ground state of the Mn^{3+} ions is degenerate, their energy is lowered by a spontaneous distortion of the surrounding lattice—the Jahn–Teller effect⁴. The charge carriers in these materials are strongly coupled to (and mediate the ferromagnetic interaction between) the manganese ions⁵, suggesting that localized lattice distortions could also play an important role in determining the electronic and magnetic properties of these compounds. Here we investigate this possibility by examining the effect on the ferromagnetic transition temperature of varying the oxygen isotope mass (replacing ^{16}O with ^{18}O). For $\text{La}_{0.8}\text{Ca}_{0.2}\text{MnO}_{3-y}$, we measure an isotope shift of $>20 \text{ K}$, significantly larger than that found for any magnetic or electronic phase transition in other oxides. In contrast, we observe no significant isotope shift for the structurally related ferromagnet SrRuO_3 , in which the Jahn–Teller

effect is negligible. These results imply that the large isotope shift arises from coupling of the charge carriers to Jahn–Teller lattice distortions, and we suggest that such Jahn–Teller ‘polarons’ may also be responsible for the magnetoresistive properties of these materials.

In 1983, Höck *et al.*⁶ studied Jahn–Teller (JT) ions in a conductor within a linear chain model. They showed that a small JT polaron (the combination of the electron and induced lattice distortion around it) can be formed when the JT stabilization energy is comparable with the bare conduction bandwidth. This is due to the breakdown of the Born–Oppenheimer adiabatic approximation in which the electronic and lattice subsystems are decoupled. The JT polaron is induced by the electronic ground-state degeneracy rather than by a charge polarization that may induce a dielectric polaron.

The polaronic nature of the conduction carriers can be demonstrated by the isotope effect on the effective bandwidth W_{eff} of polarons, which in turn depends on the isotope mass⁷:

$$W_{\text{eff}} \propto W \exp(-\gamma E_b/\hbar\omega) \quad (1)$$

Here W is the bare conduction bandwidth, E_b is the binding energy of the polaron (independent of the isotope mass), ω is the characteristic frequency of the optical phonons depending on the isotope mass M ($\omega \propto M^{-1/2}$). The dimensionless parameter γ is a function of E_b/W with $0 < \gamma \leq 1$; as E_b/W decreases, γ decreases and W_{eff} increases. For JT polarons, the binding energy of polarons E_b can be replaced by the JT stabilization energy E_{JT} (ref. 8). As the JT state of Mn^{3+} has a sizeable E_{JT} of $\sim 0.5 \text{ eV}$ (ref. 9), one would expect from equation (1) that the isotope effect on W_{eff} is substantial in the manganites.

In the ferromagnetic manganites such as $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3+y}$, it is

easy to study the isotope effect on W_{eff} . This is due to the fact that in the strong-coupling limit, where the Hund's-rule coupling J_H is very large compared to W_{eff} ($J_H \gg W_{\text{eff}}$), the Curie temperature $T_c \propto W_{\text{eff}}$ (refs 5,9). The isotope dependence of W_{eff} can also be investigated in a ferromagnet as SrRuO_3 , where T_c depends strongly on W_{eff} , as indicated by a large pressure effect¹⁰. In the ferromagnet SrRuO_3 , however, the JT effect is very weak, so the isotope dependence of W_{eff} might be very weak. Therefore, studies of the oxygen isotope effect on the Curie temperature in these ferromagnets can assess the importance of the JT effect in the formation of polarons. This will also give an important insight into the microscopic origin of the colossal magnetoresistance (CMR) effect.

Samples of $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3+y}$ were prepared by conventional solid-state reaction using La_2O_3 , CaCO_3 and MnO_2 . The powders were mixed, ground thoroughly and heated in air at $1,050^\circ\text{C}$ for ~ 20 h. The samples were then pressed into pellets, and sintered in air at $1,080^\circ\text{C}$ for ~ 48 h. Similar conditions were used for preparing samples of SrRuO_3 . The samples are single-phase, as checked by X-ray diffraction. The ^{16}O and ^{18}O samples were prepared from the same batch of the starting material, and were subjected to the same thermal treatment in closed ampoules (one was filled with $^{16}\text{O}_2$ gas, and another with $^{18}\text{O}_2$ gas)¹¹. The diffusion was carried out for 48 h at 950°C and in an oxygen pressure of ~ 1.0 bar. The difference between the oxygen partial pressures of the ^{16}O and ^{18}O ampoules was less than 4%. The cooling was very slow, with a rate of 30°C h^{-1} . The oxygen-isotope enrichment was determined from the weight changes of both ^{16}O and ^{18}O samples. The ^{18}O samples of $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3+y}$ had $\sim 95\%$ ^{18}O and $\sim 5\%$ ^{16}O . The ^{18}O samples of SrRuO_3 had $80 \pm 15\%$ ^{18}O and $20 \pm 15\%$ ^{16}O .

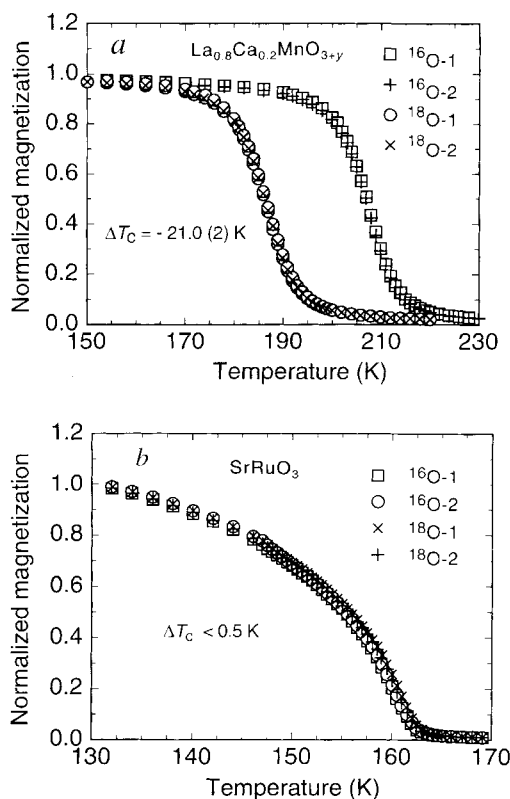


FIG. 1 Oxygen isotope effect on the Curie temperature of $\text{La}_{0.8}\text{Ca}_{0.2}\text{MnO}_{3+y}$ and SrRuO_3 : temperature dependence of the normalized magnetization for ^{16}O and ^{18}O samples of $\text{La}_{0.8}\text{Ca}_{0.2}\text{MnO}_{3+y}$ (a) and SrRuO_3 (b). There is a large oxygen isotope effect on Curie temperature (~ 21 K) in $\text{La}_{0.8}\text{Ca}_{0.2}\text{MnO}_{3+y}$ with a strong JT effect, but no observable effect in the ferromagnet SrRuO_3 with a negligible JT effect. Note that the two independent ^{16}O samples (as well as the two independent ^{18}O samples) have the same T_c , demonstrating an excellent reproducibility of the isotope experiments.

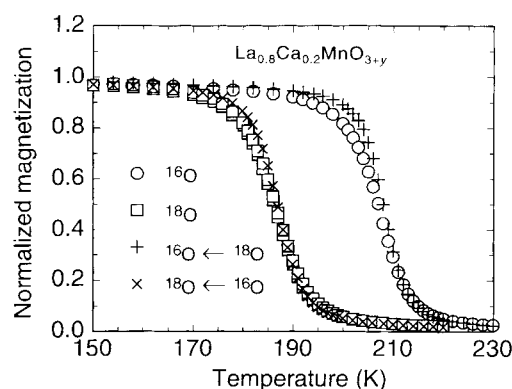


FIG. 2 Oxygen back-exchange result for sample pair ^{16}O -1/ ^{18}O -1 (see Fig. 1a): temperature dependence of the normalized magnetization for ^{16}O and ^{18}O samples of $\text{La}_{0.8}\text{Ca}_{0.2}\text{MnO}_{3+y}$ before and after oxygen back-exchange. The symbol '+' denotes the ^{16}O sample which has been back-exchanged from the original ^{18}O sample (denoted by open square). The symbol 'x' represents the ^{18}O sample which has been back-exchanged from the original ^{16}O sample (denoted by open circle). It is evident that the T_c of the ^{16}O (^{18}O) sample goes back completely to that of the original ^{18}O (^{16}O) sample after the oxygen back-exchange.

The field-cooled magnetization of the samples was measured with a commercial SQUID magnetometer in a field of 5 mT. The samples were cooled directly to 5 K, then warmed to a temperature well below T_c and maintained at that temperature for 20 minutes. Data were collected on warming to a temperature well above T_c . The magnetic field was kept unchanged throughout each series of measurements. In Fig. 1 we show the magnetization (normalized to the magnetization well below T_c) for pairs of ^{16}O and ^{18}O samples (that is, two for each isotope) of $\text{La}_{0.8}\text{Ca}_{0.2}\text{MnO}_{3+y}$ (Fig. 1a) and SrRuO_3 (Fig. 1b). The oxygen isotope shifts of T_c were determined from the differences between the midpoint temperatures on the transition curves of the ^{16}O and ^{18}O samples. For the CMR ferromagnet $\text{La}_{0.8}\text{Ca}_{0.2}\text{MnO}_{3+y}$, the ^{18}O samples have lower T_c s than the ^{16}O samples by ~ 21 K, whereas the two ^{16}O samples (as well as the two ^{18}O samples) have the same T_c . On the other hand, no oxygen isotope effect on T_c could be detected in the ferromagnet SrRuO_3 with a negligible JT effect. If we define the oxygen isotope exponent as $\alpha_o = -\text{dln}T_c/\text{dln}M_o$ (where M_o is the oxygen isotope mass), then we obtain $\alpha_o \approx 0.85$ for $\text{La}_{0.8}\text{Ca}_{0.2}\text{MnO}_{3+y}$. For $\text{La}_{0.9}\text{Ca}_{0.1}\text{MnO}_{3+y}$, we found a slightly smaller exponent $\alpha_o \approx 0.70$ ($T_c \approx 140$ K, $\Delta T_c = 11.7(2)$ K). These exponents are considerably larger than those recently observed in the Sr-doped system $\text{La}_{1-x}\text{Sr}_x\text{MnO}_{3+y}$ (for example, $\alpha_o = 0.19$ for $x = 0.10$; $\alpha_o = 0.14$ for $x = 0.15$; $\alpha_o = 0.07$ for $x = 0.3$; G.M.Z. and D. E. Morris, unpublished results).

To show that the observed oxygen isotope shifts are intrinsic, we have performed oxygen back-exchange experiments ($^{16}\text{O} \rightarrow ^{18}\text{O}$; $^{18}\text{O} \rightarrow ^{16}\text{O}$). In Fig. 2 we show the normalized magnetization for the ^{16}O and ^{18}O samples of $\text{La}_{0.8}\text{Ca}_{0.2}\text{MnO}_{3+y}$ before and after oxygen back-exchange. It is evident that the T_c of the ^{16}O (^{18}O) sample goes back completely to that of the original ^{18}O (^{16}O) sample after the oxygen back-exchange. This clearly indicates that the shift of T_c is caused only by changing the oxygen isotope mass.

It is also important to check whether the observed isotope shifts may arise from a difference in the oxygen contents of the ^{16}O and ^{18}O samples. We think this is very unlikely for the following reasons. First, the T_c of the $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3+y}$ system is hardly changed even after some extreme thermal treatments (for example, annealing under 200 bar of oxygen pressure at $\sim 600^\circ\text{C}$ and quenching from $1,300^\circ\text{C}$)¹², and the present ^{16}O and ^{18}O samples were treated under the same thermal condition. Second, the ^{16}O and ^{18}O samples of the related $\text{La}_{2-x}\text{Sr}_x\text{CuO}_{4+y}$ compound were shown to have the same oxygen contents by comparing the lattice

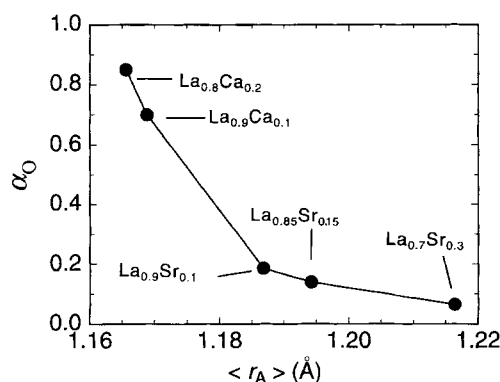


FIG. 3 The dependence of the oxygen isotope exponent α_O on the average ionic radius $\langle r_A \rangle$ at the cation site $\text{La}_{1-x}\text{Me}_x$. (The results for the Sr-doped compounds are unpublished results of G.M.Z. and D. E. Morris.) The solid line is to guide the eye. The isotope exponent α_O increases rapidly with decreasing $\langle r_A \rangle$.

constants of the two isotope samples¹³. For the present CMR materials, we have used the same oxygen-isotope exchange procedure as for the $\text{La}_{2-x}\text{Sr}_x\text{CuO}_{4+y}$ compound, so we expect that the oxygen contents of the present ^{16}O and ^{18}O samples are the same as well.

We now consider why the oxygen isotope exponents α_O in Ca-doped $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3+y}$ are much larger than those in Sr-doped $\text{La}_{1-x}\text{Sr}_x\text{MnO}_{3+y}$. This may be due to the difference in the ionic radii of Ca^{2+} and Sr^{2+} . In Fig. 3 we show the dependence of α_O on the average ionic radius $\langle r_A \rangle$ at the cation site $\text{La}_{1-x}\text{Me}_x$. The ionic radii of $\text{La}_{1-x}\text{Me}_x$ were calculated from tabulated values (La^{3+} , 1.172 Å; Ca^{2+} , 1.140 Å; Sr^{2+} , 1.320 Å; ref. 14). It is evident that α_O increases rapidly with decreasing $\langle r_A \rangle$. Although α_O may also depend on the Mn^{4+} concentrations, it is clear from Fig. 3 that $\langle r_A \rangle$ has a dominant effect on the value of α_O . A similar correlation between the magnitude of the magnetoresistance and $\langle r_A \rangle$ has recently been reported by Hwang *et al.*¹⁵; they showed that the magnetoresistance decreases rapidly with increasing $\langle r_A \rangle$. Thus, the oxygen-isotope exponent α_O is correlated with the magnitude of magnetoresistance.

As there is no oxygen-isotope effect on T_c in the ferromagnet SrRuO_3 which has a negligible JT effect, the giant isotope effect observed in the Ca-doped ferromagnets is very likely to be related to the strong JT effect in this system. For a compound with a strong JT effect, the electron-phonon interaction is also large, leading to the formation of JT polarons⁶. Taking into account that the binding energy of polarons E_b in the JT compound is equal to the JT stabilization energy E_{JT} (ref. 8), and $T_c \propto W_{\text{eff}}$ for $J_H \gg W_{\text{eff}}$, equation (1) gives

$$T_c \propto W \exp(-\gamma E_{JT}/\hbar\omega) \quad (2)$$

The total isotope exponent is then given by

$$\alpha = -d\ln T_c/d\ln M = 0.5\gamma E_{JT}/\hbar\omega \quad (3)$$

Because γ increases with increasing E_{JT}/W (ref. 7) equation (3) indicates that the isotope exponent increases with increasing E_{JT} , but decreases with increasing W . For the manganites, an increase in $\langle r_A \rangle$ usually enhances the covalency of the Mn-O bonding¹⁶, and hence increases the bare conduction bandwidth W , leading to a decrease of α . This mechanism naturally explains why the oxygen-isotope exponent α_O increases with decreasing $\langle r_A \rangle$, as shown in Fig. 3.

Our results thus strongly suggest that in the CMR manganites there is a substantial JT effect and thus a large electron-phonon interaction which leads to the formation of JT polarons⁶. This is also consistent with a recent theoretical study of the JT lattice coupling effects in the CMR manganites¹⁷. For the ferromagnet SrRuO_3 with a negligible JT effect, the electron-phonon interaction

may be very small, so that no polarons are formed. That is why no oxygen isotope effect on the Curie temperature is observed in the ferromagnet SrRuO_3 . □

Received 8 March; accepted 9 May 1996.

1. Jin, S. *et al.* *Science* **264**, 413–415 (1994).
2. Jonker, G. H. & Van Santen, J. H. *Physica* **16**, 337–349 (1950).
3. Chahara, K. *et al.* *Appl. Phys. Lett.* **63**, 1990–1992 (1993). (see note below)
4. Jahn, H. A. & Teller, E. *Proc. R. Soc. Lond. A* **161**, 220–235 (1937).
5. Anderson, P. W. & Hasegawa, H. *Phys. Rev.* **100**, 675–681 (1955).
6. Höck, K.-H., Nickisch, H. & Thomas, H. *Helv. phys. Acta* **50**, 237–243 (1983).
7. Alexandrov, A. S. & Mott, N. F. *Int. J. mod. Phys. B* **8**, 2075–2109 (1994).
8. De Jongh, L. J. *Physica C* **152**, 171–216 (1988).
9. Millis, A. J., Littlewood, P. B. & Shraiman, B. I. *Phys. Rev. Lett.* **74**, 5144–5147 (1995).
10. Shikano, M. *et al.* *Solid St. Commun.* **90**, 115–119 (1994).
11. Zech, D. *et al.* *Nature* **371**, 681–683 (1994).
12. Schiffer, P. *et al.* *Phys. Rev. Lett.* **75**, 3336–3339 (1995).
13. Zhao, G. M. *et al.* *Phys. Rev. B* **52**, 6840–6844 (1995).
14. Shannon, R. D. *Acta crystallogr.* **A32**, 751–767 (1976).
15. Hwang, H. Y. *et al.* *Phys. Rev. Lett.* **75**, 914–917 (1995).
16. Goodenough, J. B. *Phys. Rev.* **106**, 564–573 (1955).
17. Röder, H., Zhang, J. & Bishop, A. R. *Phys. Rev. Lett.* **76**, 1356–1359 (1996).

ACKNOWLEDGEMENTS. This work was supported by the Swiss National Science Foundation.

Exceptionally high Young's modulus observed for individual carbon nanotubes

M. M. J. Treacy*, T. W. Ebbesen* & J. M. Gibson†

*NEC Research Institute, Inc., 4 Independence Way, Princeton, New Jersey 08540, USA

†University of Illinois, Department of Physics, 1110 West Green Street, Urbana, Illinois 61801, USA

CARBON nanotubes are predicted to have interesting mechanical properties—in particular, high stiffness and axial strength—as a result of their seamless cylindrical graphitic structure^{1–5}. Their mechanical properties have so far eluded direct measurement, however, because of the very small dimensions of nanotubes. Here we estimate the Young's modulus of isolated nanotubes by measuring, in the transmission electron microscope, the amplitude of their intrinsic thermal vibrations. We find that carbon nanotubes have exceptionally high Young's moduli, in the terapascal (TPa) range. Their high stiffness, coupled with their low density, implies that nanotubes might be useful as nanoscale fibres in strong, lightweight composite materials.

Large bundles of carbon nanotubes were teased from a core

TABLE 1 Properties of individual nanotubes

Nanotube no.	Length (μm)	Outer diameter (nm)	Inner diameter (nm)	Young's modulus (TPa)
1	1.17	5.6	2.3	1.06
2	3.11	7.3	2.0	0.91
3	5.81	24.8	6.6	0.59
4	2.65	11.9	2.0	1.06
5	1.73	7.0	2.3	2.58
6	1.53	6.6	2.3	3.11
7	2.04	7.0	3.0	1.91
8	1.43	6.6	3.3	4.15
9	0.66	7.0	3.3	0.42
10	1.32	9.9	3.0	0.40
11	5.10	8.4	1.0	3.70

Average value of Young's modulus is 1.8 TPa.

deposit, which had been prepared by the carbon arc method^{1,6}, and attached to the edge of a hole in 3-mm-diameter nickel rings for transmission electron microscopy (TEM) observations. The tip of the fibrous bundle was gently shredded by pulling with tweezers to produce a brush of roughly horizontal isolated nanotubes, which are effectively clamped at one end. The specimens were examined in a Hitachi H9000NAR TEM at 100 kV, and were supported on a Gatan 652-Ta side-entry heating holder equipped with a thermocouple. A Gatan model 690 slow-scan CCD (charge-coupled device) camera, in conjunction with Digital Micrograph v2.5 software, was used to collect and process image data.

In the TEM at room temperature, it was found to be impossible to focus on the tips of the longest and thinnest nanotubes. The images were significantly blurred at the free ends, but could be brought sharply into focus at the supported ends. Refocusing demonstrated that the tip blurring was not caused by focal gradients due to nanotubes being inclined to the optic axis. Increasing the specimen temperature to 600 K significantly increased the blurring at the tips. This result indicates that these nanotubes are vibrating, and that the vibration is of thermal origin. Figure 1 compares the bright-field images of several micrometre-long nanotubes at 300 and 600 K. Near the supported base they measure ~ 15 nm across. At the tips, however, image contrast is lower. At 300 K the imaged tip width of one of them is ~ 30 nm, whereas at 600 K the width is ~ 40 nm. Decreasing the electron beam current density by a factor of 1,000 does not noticeably affect the motion. Detailed measurements of the vibration amplitude with temperature (Fig. 2) confirm the thermal origin. Occasionally, nanotubes are observed with relatively massive lumps attached near their end (see, for example, the nanotube on the right of each panel in Fig. 1). These weighted nanotubes sway and twitch laterally in a random manner that is suggestive of brownian motion.

We assume that a nanotube is equivalent to a clamped homogeneous cylindrical cantilever of length L , with outer and inner radii a and b , respectively. It can be shown that for a given mode, n , the horizontal vibration amplitude at the tip, u_n , is related to the Young's modulus Y and the vibration energy W_n through⁷

$$W_n = \frac{1}{2} c_n u_n^2 \quad (1)$$

where the effective spring constant for the motion of the tip is

$$c_n = \frac{\pi \beta_n^4 Y (a^4 - b^4)}{16 L^3} \quad (2)$$

The values of β_n are found from the solutions to the equation $\cos \beta \cosh \beta + 1 = 0$. $\beta_0 \approx 1.8751$ for the fundamental mode, and $\beta_1 \approx 4.6941$, $\beta_2 \approx 7.8548$, $\beta_3 \approx 10.9955$ for the first three overtones. $\beta_n \approx (n + 1/2)\pi$ for larger n . The undamped vibration frequency is³

$$\omega_n = 2\pi f_n = \frac{\beta_n^2}{2L^2} \sqrt{\frac{Y(a^2 + b^2)}{\rho}} \quad (3)$$

where ρ is the density of the material in the nanotube wall. For typical nanotubes at room temperature, we expect $\omega_0 \approx 10^7$ Hz, and $kT \gg \hbar \omega_n$ for $n \leq 1,000$. Consequently, from the law of equipartition, there is an average energy of $kT/2$ per degree of freedom for all of the relevant lateral vibration modes. Because there are both elastic and kinetic energy degrees of freedom for all of the relevant lateral vibration modes. Because there are both elastic and kinetic energy degrees of freedom in a vibrational mode then, on average, $\langle W_n \rangle = kT$ for each vibration mode, with W_n obeying the Boltzmann distribution. It is straight-

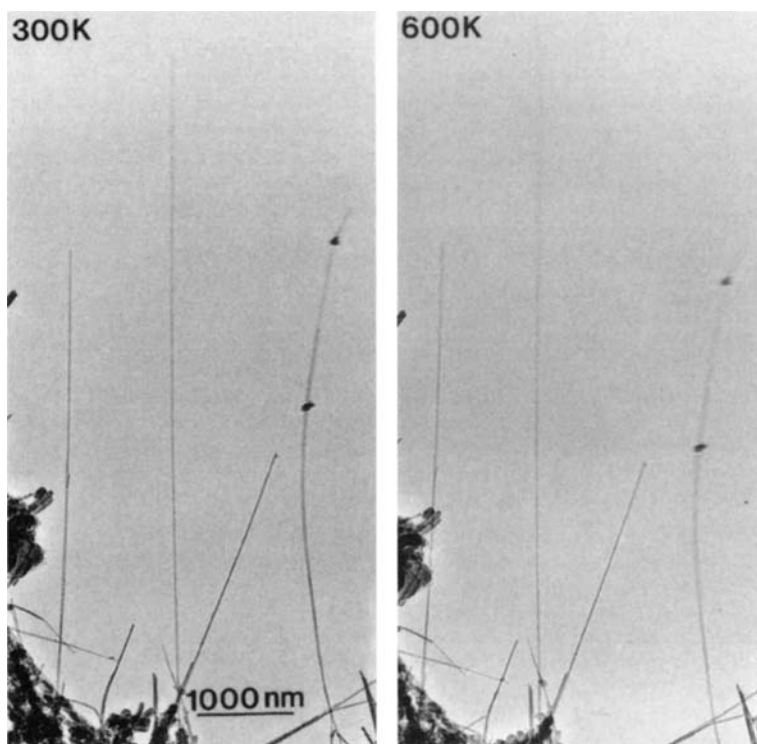


FIG. 1 Bright-field TEM micrographs of free-standing carbon nanotube fibres showing the blurring at the tips due to thermal vibration. Left panel, 300 K; right panel, 600 K.

forward to show that each mode of a stochastically-driven oscillator has a gaussian vibration probability profile, with standard deviation given by $\sigma_n = (kT/c_n)^{1/2}$. As the modes are mutually independent, the vibration profile for the combined modes is also a gaussian, with standard deviation given by the

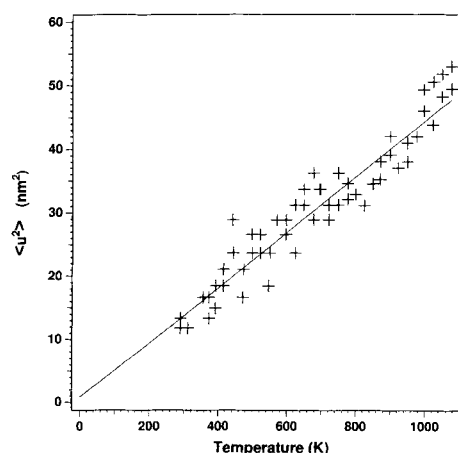


FIG. 2 Plot of the mean-square vibration amplitude versus temperature for a 5.1- μ m-long, 16.6-nm-wide, free-standing nanotube. The slope is $0.044 \pm 0.0018 \text{ nm}^2 \text{ K}^{-1}$, giving an effective Young's modulus of $3.7 \pm 0.2 \text{ TPa}$. The nanotube was heated from room temperature to 800 °C in steps of 25 °C. At each temperature, several low-magnification bright-field images of the vibrating tip were acquired on the slow-scan CCD camera at a resolution of 1.65 nm pixel. To improve contrast, a small objective aperture of semi-radius 0.7 nm⁻¹ was employed. The small aperture also increases the depth of focus, thereby minimizing the additional tip image blurring due to the vertical vibration of the nanotube. Images were collected with $\sim 1,000$ counts per pixel. Line traces were averaged over a 40 nm length at the nanotube tip. The room-temperature image profile of the base of the nanotube was convoluted with a gaussian to match the line trace of the tip.

sum of the variances

$$\sigma^2 = \frac{16L^3kT}{\pi Y(a^4 - b^4)} \sum_n \beta_n^{-4} \approx 0.4243 \frac{L^3kT}{Y(a^4 - b^4)} \quad (4)$$

If we assume $a = 8$ nm, $b = 1$ nm, $L = 5,000$ nm and $Y = 5 \times 10^{11}$ Pa (which is typical for macroscopic carbon fibres⁸), then equation (4) gives a root-mean-square vibration amplitude of ~ 10 nm at room temperature, which agrees qualitatively with the observed blurring at the tip. An equivalent vibrational motion is occurring in the vertical direction, but we could not detect this motion in the TEM. In principle, there are also thermal longitudinal oscillations, and torsional modes. The mean square longitudinal thermal vibration amplitude is found from $\langle z^2 \rangle = LkT/\pi Y(a^2 - b^2)$, giving $\langle z^2 \rangle^{1/2} \approx 0.007$ nm, which is negligible. The fundamental torsional mode generates a mean square twist $\langle \theta^2 \rangle = LKT/2\mu\pi(a^4 - b^4)$. Assuming that the shear modulus $\mu = Y/2(1 + \sigma)$, and that Poisson's ratio $\sigma \approx 0.3$, we find that $\langle \theta^2 \rangle^{1/2}$ is ~ 2 mrad. From equations (1)–(3), and the law of equipartition of energy, we expect each mode of the 'noisy' tip motion to follow $\langle u_n^2 \rangle^{1/2} \propto 1/f_n$ for an individual nanotube, with the constant of proportionality governed by $[L(a^2 - b^2)\rho]^{-1/2}$.

Because of the low image contrast at the tip, and the absence of the normal Fresnel fringes, great care is necessary to ensure that the nanotube tips are at the correct focus. Because the longitudinal vibration amplitudes at the end of the nanotube are small, spatial frequencies parallel to the fibre axis are not blurred. Consequently, Fresnel fringes at the top of the nanotube can be used to focus the vibrating tip. By measuring the change in focus between the base of the nanotube and the tip, it is possible to deduce the inclination of the nanotube to the horizontal. This assumption is reasonable because the droop of the nanotube under the influence of gravity, $u_g = L^4\rho g/2Y(a^2 + b^2)$, is less than 3×10^{-4} nm if we assume $\rho = 2,150$ kg m⁻³, and so the effect of gravity may be ignored.

The width of the tip image is essentially the convolution of the image of the relatively stationary nanotube base with the vibration profile. If we assume that the image profile of the (stationary) tip is similar to that at the base, then the vibration amplitude is obtained, in principle, by deconvoluting the base image profile from the tip image profile. In practice, as we know the form of the vibration profile, better results are obtained by convoluting the base profile with a gaussian and matching the result to the tip profile. For a given vibrating nanotube, the assumption of structural uniformity can be tested by comparing the total scattered intensity under the blurred tip image profile with that of the base. When the image focus is the same in each case, then the total scattering, as estimated by the area under the bright-field image profiles, should be identical. Discrepancies can be used to distinguish fibres with varying diameter along the length. For the nanotubes studied here, if the area under the scattering profile of the tip deviated by more than $\pm 10\%$ of that at the base, the data was rejected. In this study, we did not attempt to measure tip vibration frequencies. Spectral analysis of the undamped cantilever modes, which are expected to be among the lowest vibration

frequencies of the nanotube, would provide an independent measure of Y through equation (3).

Figure 2 shows a graph of the measured mean-square vibration amplitude $\langle u^2 \rangle$ versus temperature T for a single nanotube. A good straight-line fit is obtained which yields a Young's modulus of $(3.7 \pm 0.2) \times 10^{12}$ Pa, which is about a factor of five higher than that measured for carbon fibres⁵. The intercept on the temperature axis occurs at $T_p \approx -13 \pm 29$ K. This value implies that the electron beam, and other non-equilibrium perturbations, contribute $\approx +13$ K of additional heating, and confirms that non-equilibrium perturbations are small. Table 1 summarizes the results for a variety of nanotubes of different inner and outer diameters. We obtain an average value for Young's modulus, $Y = 1.8$ TPa. Note that, on average, the measured modulus is highest for the thinner nanotubes. There is a full order-of-magnitude range in the values for Y . The spread in values stems partially from inevitable experimental uncertainties, such as in the estimation of the length of the tubes from their true point of anchoring, but most importantly from variation in structural features away from perfectly nested cylinders. For instance, should the nanotube have any inner ('bamboo-like') compartments with only the outer few layers being continuous from one end to the other, the hinging action at the joint will cause the modulus to be underestimated when we use equation (4). However, all these errors will tend to underestimate the flexural rigidity, or stiffness. We expect the correct modulus to tend towards the higher values.

It should be noted that the nanotube dimensions are optimal for direct observation of thermal vibrations. Therefore, the mechanical properties of other types of nanoscopic fibres, which are inaccessible to traditional techniques, could be studied by this method.

The values for the Young's modulus we find are much higher than those for typical carbon fibres⁵. For instance, an average value of 6.8×10^{11} Pa was obtained for macroscopic vapour-grown carbon fibres⁸. Carbon whiskers made by Bacon⁹ had, so far, the highest measured extensional moduli of ~ 800 GPa. The in-plane modulus of single crystal graphite is of the order of 1 TPa. Therefore, by wrapping the graphite sheets into seamless cylinders, nanotubes appear to develop more stiffness than the graphite in-plane modulus would imply. Theoretical estimates predict in one case² that single-walled nanotubes might have a Young's modulus of the order of 5 TPa. Others, however, predict a softening of the nanotube with decreasing radius and show a distinct dependence on the helical pitch, the upper value being limited by that of planar graphite³. These studies neglect inter-layer interactions which might contribute to a stiffening of the nanotubes.

Carbon fibres are already widely used in reinforced composites where high-strength light-weight materials are needed. As pointed out by Calvert⁴, there is need for small-diameter, large-aspect-ratio fibres to obtain even higher-strength composites. Further, small-fibre composites should be easier to process⁴. In this context, the exceptional mechanical properties of nanotubes, combined with their small dimensions and large aspect ratio, could lead to optimal carbon-fibre-reinforced materials. $\square$

Received 18 March; accepted 14 May 1996.

1. Ebbesen, T. W. A. *Rev. Mater. Sci.* **24**, 235–264 (1994).
2. Overney, G., Zhong, W. & Tomanek, D. Z. *Phys.* **D27**, 93–96 (1993).
3. Robertson, D. H., Brenner, D. W. & Mintmire, J. W. *Phys. Rev.* **B45**, 12592–12595 (1992).
4. Calvert, P. *Nature* **357**, 365–366 (1992).
5. Dresselhaus, M. S., Dresselhaus, G., Sugihara, K., Spain, I. L. & Goldberg, H. A. in *Graphite Fibers and Filaments* Ch. 6 (eds Gonser, U., Mooradian, A., Muller, K. A., Panish, M. B. &

- Sakaki, H.) 120–152 (Springer, New York, 1988).
6. Ebbesen, T. W. & Ajayan, P. M. *Nature* **358**, 220–222 (1992).
7. Kreyszig, E. *Advanced Engineering Mathematics* 7th edn. 643–644 (Wiley, New York, 1993).
8. Jacobsen, R. L., Tritt, T. M., Guth, J. R., Ehrlich, A. C. & Gillespie, D. J. *Carbon* **33**, 1217–1221 (1995).
9. Bacon, R. J. *J. appl. Phys.* **31**, 283–290 (1960).

ACKNOWLEDGEMENTS. We wish to thank M. E. Bisher for valuable assistance.

Direct radiative forcing by anthropogenic airborne mineral aerosols

Irina N. Sokolik* & Owen B. Toon

Earth Science Division, Ames Research Center, National Aeronautical and Space Administration, Mail Stop 245-4, Moffett Field, California 94035, USA

AIRBORNE mineral dust can have a significant effect on the Earth's radiation budget, as it can both scatter sunlight back to space (leading to negative radiative forcing), and absorb solar and infrared radiation (leading to positive forcing)^{1,2}. The effects of mineral aerosols on the radiation budget are important relative to those of other types of aerosols—such as sulphate and smoke particles—due to the widespread distribution and large optical depth of mineral dust. Various human activities, such as land use practices, can result in additional loading of dust, increasing the radiative forcing. Previous studies have attempted to estimate the radiative effects of both the natural and anthropogenic components of the dust^{3,4}. Here we use estimates of anthropogenic dust inputs and observations of dust optical properties to show that although the key quantities contributing to the evaluation of the direct solar radiative forcing by dust generated through human activities have a wide range of uncertainty, the forcing by anthropogenically generated mineral aerosols may be comparable to the forcing by other anthropogenic aerosols. On a regional scale the forcing due to mineral aerosols can greatly exceed that due to sulphate aerosols and can be comparable to that of clouds. Our analysis enables us to highlight the key quantities that need to be better characterized to reduce the (currently large) uncertainties in these estimates.

The direct solar radiative forcing by anthropogenic sulphate aerosols^{5,6} and smoke particles from biomass burning^{7,8} have been much studied, but the forcing by anthropogenic mineral dust has received far less attention^{1-4,9}. Tegen and Fung³ concluded that surface land modification by human activities could account for 30–50% of the total atmospheric dust loading, and very recently Tegen *et al.*⁴ suggested that inputs of this magnitude could decrease the net global surface radiative forcing by around 1 W m⁻². The uncertainties in such estimates remain substantial.

To assess the direct radiative forcing of airborne dust in comparison to that produced by anthropogenic sulphate aerosols or aerosols produced by biomass burning, we have employed the approach described in refs 5–8. The globally averaged direct solar radiative forcing, ΔF , for an optically thin layer of absorbing aerosols can be written in the form

$$\Delta F = -K_{\text{atm}} K_{\text{opt}} \tau$$

where the term $K_{\text{atm}} = (S_0/4)T^2(1-N)$ does not depend on dust characteristics. S_0 is the solar constant, T is the atmospheric transmission above the aerosol layer, N is the cloud fraction. Adopting $S_0 = 1,370 \text{ W m}^{-2}$, $T = 0.79$ and $N = 0.6$ (refs 7,8), we find K_{atm} is 85.5 W m^{-2} . In turn, $K_{\text{opt}} = \omega(1-g/2)(1-a)^2 - 4a(1-\omega)$ depends on dust optical properties and globally averaged surface albedo, a . Here ω is the aerosol single scattering albedo, g is the asymmetry parameter, and τ is the global-mean dust optical depth.

The coefficient K_{opt} can be estimated from available dust optical models. Figure 1a, b presents spectra of ω and g for published dust models¹⁰⁻¹⁵. These models give a wide range for K_{opt} , namely -0.08 to 0.35 over the land and 0.24 – 0.55 over the ocean. The positive K_{opt} implies negative radiative forcing, and negative K_{opt} indicates warming. Therefore, a determination of the sign of ΔF requires better knowledge of the optical properties of mineral aerosols.

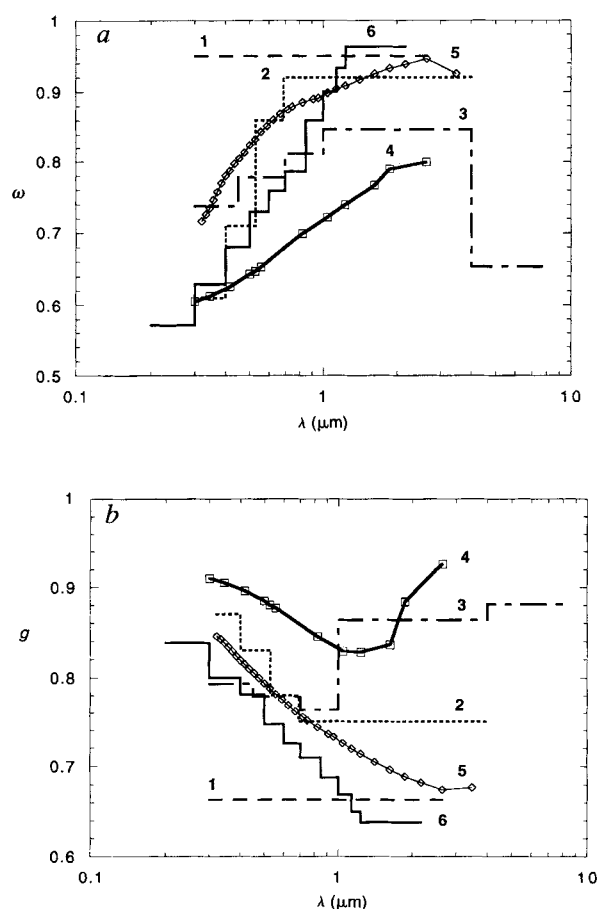


FIG. 1 Spectral dependence of the single scattering albedo ω (a) and asymmetry parameter g (b) from optical dust models: curve 1 is for the model of Fouquart *et al.*¹⁰; curve 2 is for Carlson and Benjamin¹¹; curve 3 is for D'Almeida¹²; curve 4 is for WMO¹³; curve 5 is for Sokolik and Golitsyn¹⁴; and curve 6 is for Ackerman and Cox¹⁵. The discrepancies between the various dust optical models are due both to differences in selected particle size distributions between the models and differences in prescribed refractive indices between the models^{14,29}. These discrepancies are caused by inability to model the physical processes responsible for dust optical properties, measurement uncertainties, and also by actual differences in properties of mineral aerosols originating from various geographical regions, which can be a serious complication in assessment of the regional dust effects. Unfortunately there is no way to decide which of the various models is more appropriate.

To estimate the global-mean anthropogenic dust optical depth we examine the major dust production mechanisms. Although there are many examples of humans altering their environment and thereby causing an additional dust burden, it is a challenging problem to quantify separately the natural and anthropogenic components of the mineral aerosol¹⁶. In recent decades the dramatic increase in the human population has had a large effect on Earth's drylands, causing degradation. Recognizing the severity of the problem, UNEP¹⁷ introduced the term 'desertification' for land degradation resulting mainly from adverse human effects. At present, desertification threatens about one-third of the world's land surface, spreading over every continent with an estimated rate¹⁷ of about $6 \times 10^4 \text{ km yr}^{-1}$. There are many man-made factors involved in desertification which result in additional loading of mineral particles into the atmosphere. Some major factors are: over-cultivation of poor soils, inappropriate irrigation practices, human-induced wind erosion, severe trampling and heavy grazing, deforestation, urbanization (for example, building and road construction), mining and industry, off-road vehicle use and tourism¹⁷.

Estimates of dust production from the global dust sources at

* To whom correspondence should be addressed.

their modern geographical boundaries cover the range between 500 and 5,000 Tgyr⁻¹, with most likely dust production rates of a few thousand Tgyr⁻¹ (refs 18–20).

For a given dust production rate, R , the global-mean optical depth is

$$\tau = R t \alpha / A$$

where t is dust lifetime, A is Earth's surface area (5.1×10^{14} m²) and α is the mass extinction coefficient of the mineral aerosols. For dust with optically significant particle sizes, which has been transported for several hundred kilometres, the typical residence time is a few days²⁰. The mass extinction coefficient can vary in a range of 0.2–2 m² g⁻¹, depending mainly on dust refractive indices²¹. Taking for simplicity $t = 4$ days and $\alpha = 1$ m² g⁻¹, we estimate τ in the range 0.017–0.085 for dust production rates $R = 1,000$ –5,000 Tgyr⁻¹.

By assuming that the dust production rate is linearly proportional to the dust source area, we can roughly estimate the amount of anthropogenic mineral aerosols through assessments of the land area converted to deserts by human activities. UNEP¹⁷ estimated $(10$ – $16) \times 10^6$ km² of drylands have been converted to desert by human-induced soil degradation. Currently, the total area of dust sources are about 30% of the continental surface or about 50×10^6 km². Therefore anthropogenic dust is about 20–30% of the total dust loading. A recent independent study³ gives an area of about 7×10^6 km² for dust sources induced by human activities during past 50 years. However, it was assumed in that study that such lands are much more efficient dust sources than natural deserts, so 30–50% of the current dust optical depth was ascribed to anthropogenic sources.

Taking anthropogenic dust loading to be 20–30% of the total, we estimate that a global-mean optical depth of 0.01–0.016 (for a total production rate of 3,000 Tgyr⁻¹) may be of anthropogenic origin. These optical depths (which may underestimate anthropogenic dust production because they do not include anthropogenic dust generation in natural deserts or the greater dust generation efficiency in newly created arid lands) are comparable to those estimated for anthropogenic sulphates, which are in the range 0.017–0.03 (refs 1,5).

Table 1 gives the most important quantities contributing to the

TABLE 1 Contributions to solar radiative forcing

Quantity	Representative value	Uncertainty range
Total dust production rate (Tgyr ⁻¹)	3,000	1,000–5,000
Anthropogenic dust production as fraction of total dust production	20%	10–50%
Lifetime of the mineral aerosols (d)	4	3–7
Mass extinction coefficient of the mineral aerosols (m ² g ⁻¹)	1	0.2–2
Estimated global-mean optical depth of anthropogenic mineral aerosols	0.014	0.003–0.05
Single scattering albedo of mineral aerosols at 500 nm	0.85	0.65–0.95
Asymmetry parameter of mineral aerosols at 500 nm	0.75	0.65–0.88
Estimated K_{opt}		
Over land	0.19	–0.08–0.35
Over ocean	0.43	0.24–0.55
Cloud fraction N	0.6	0.56–0.65
Atmospheric transmission above the aerosol layer	0.79	0.75–0.85
Estimated K_{atm} (W m ⁻²)	85.5	84.8–86.6
Estimated ΔF due to anthropogenic dust (W m ⁻²)		
Over land	–0.25	–0.08 to –0.9*
Over ocean	–0.6	–0.2 to –2.2
Estimated ΔF due to anthropogenic sulphates (ref. 22) (W m ⁻²)	–0.6	–0.26 to –1.4
Estimated ΔF due to biomass-burning smoke (ref. 22) (W m ⁻²)	–0.8	–0.1 to –2.2

* The uncertainty in ΔF was estimated, considering the uncertainty range in optical depth.

evaluation of the global-mean direct solar radiative forcing by the mineral aerosols, and compares our estimates of global radiative forcing by dust with those by anthropogenic sulphates and biomass smoke. Table 1 is made similar to a table prepared by Penner *et al.*²² to simplify comparison between the ranges of uncertainties related to our dust case study and those for other anthropogenic aerosols. Table 1 indicates that each quantity needed to assess the radiative forcing by dust has rather a wide range of uncertainty. However, an improved knowledge on the anthropogenic fraction of dust production, and of the dust optical properties, are the first

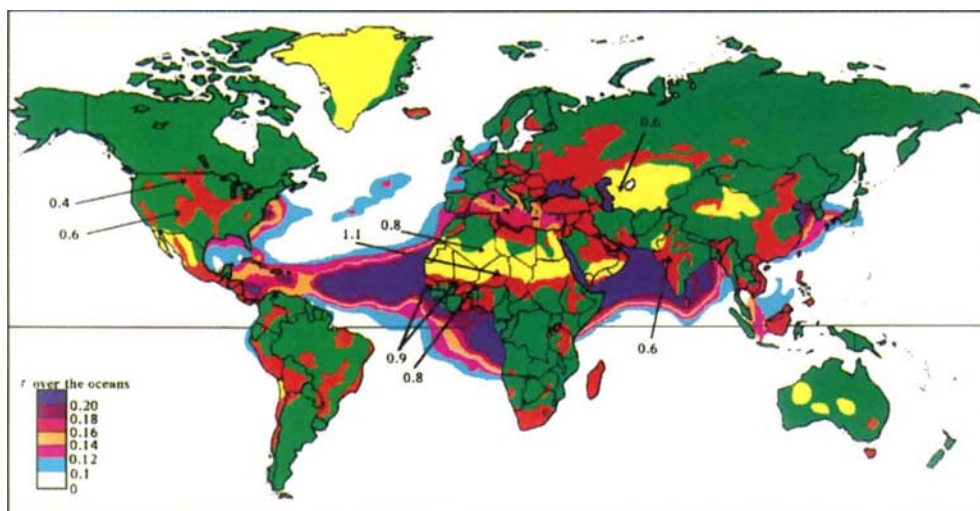


FIG. 2 Desertification map from ref. 17 with some representative aerosol optical depths in highly degraded areas. The optical depths, τ , shown as numerical values are as follows: for USA, July mean³⁰; for Africa, summer mean^{12,31}; for Russian central Asia, September mean^{32,33}; for India, July mean³⁰. Areas of very high degradation are shown in red. Optical depths over the oceans are seasonal means retrieved from satellite AVHRR data³⁴. The highest optical depths over the oceans occur in regions influenced by dust transport.

priority for better evaluation of the global-mean solar forcing by mineral aerosols.

Although the globally averaged solar radiative forcing is a useful concept, it can be used only for a first-order assessment¹. To estimate correctly the actual global forcing, the regional distribution of the radiative forcing should be considered because aerosol sources and sinks are not uniformly distributed. Figure 2 presents some ground-based and satellite measurements of dust optical depths over the severely degraded lands and over the oceans. Examination of Fig. 2 shows that there are many dispersed geographical regions where the optical depths of mineral dust are several times larger than those of sulphate aerosols and of biomass-burning smoke. Over the tropical and subtropical North Atlantic, light scattering by mineral dust is at least four times greater than that by non-seasalt sulphate⁹.

Based on observed optical depths, we deduce the possible magnitude of the regional net radiative forcing due to mineral particles using a one-dimensional radiative transfer code¹⁴. These rough estimations give a positive net (solar plus infrared) direct radiative forcing of 20–40 W m⁻² over the arid areas with season-mean τ of about 0.5 at visible wavelengths, and somewhat smaller negative forcing of -5 to -15 W m⁻² over the ocean. These estimates are supported by satellite observations from the Earth Radiation Budget Experiment (ERBE) during a dust outbreak²³. In contrast, the calculated regional direct aerosol forcing by sulphates is typically about -1 to -2 W m⁻² (though in small regions the forcing may rise to -4 to -10 W m⁻²)²⁴. The large radiative forcings of mineral aerosols near dust source areas are comparable to the forcings observed for clouds, which are in the range 10 to -10 W m⁻² from ERBE observations²⁵ in the major dust production areas, and in the larger range 20 to -20 W m⁻² in the region from 30° S to 30° N.

Unfortunately, a more adequate assessment of regional forcing by dust is hardly possible at the moment owing to the poor knowledge of dust optical properties and their time and space variations. The magnitude of direct aerosol radiative forcing depends strongly on the regional dust optical characteristics, regional surface albedo and solar zenith angle. As an illustration, Fig. 3 shows the importance of existing discrepancies in the optical models for the evaluation of the solar heating/cooling effects by mineral aerosols in various geographical regions.

Further, the net (solar plus infrared) forcing should be considered because dust can have a significant impact on thermal radiation (as we discussed above for regional forcings). Experimental data on dust optical properties at infrared wavelengths are very limited. Observations of the optical depth in the infrared window region suggest that it is about 2–10 times smaller than the visible optical depth¹⁴. Little information is available on the vertical distribution of the mineral particles, which is critical for calculation of their infrared radiative forcing.

Finally, because the airborne dust physical characteristics vary during transport because of sedimentation and wash-out, radiative forcing calculations should include the life-cycle of the mineral particles^{26–28}. We are not aware of any model that links the dust global life-cycle to its optical characteristics.

Further understanding of the radiative forcing due to dust requires an improved data base on regional dust optical properties, the anthropogenic component, the regional dust optical depth, and the evolution of dust optical properties as dust is transported downwind from the source regions. It is probable that improved dust optical properties could be obtained by more carefully relating light absorption properties to mineralogy and source region, as well as by better characterizing the evolution of dust particles with distance from the dust source region.

The above uncertainties in the dust properties do not diminish the importance of the radiative forcing by the mineral particles. The estimated global-mean solar radiative forcing by anthropogenic dust (see Table 1) is in the same range as that due to anthropogenic sulphates and biomass-burning smoke. Therefore the evolution of the dust radiative forcing over the past century

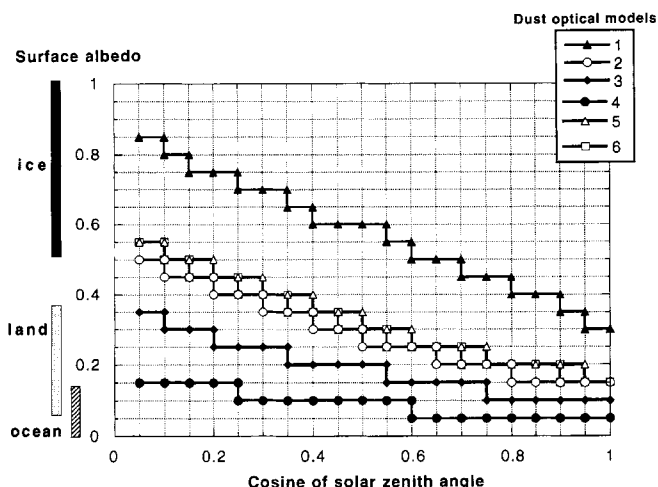


FIG. 3 Zones of solar heating/cooling of the underlying surface-atmosphere system for dust optical models (curves 1–6 in Fig. 1). In the region of the graph above each curve, the system albedo without dust loading is higher than it is with dust (for the particular model represented by the curve). Hence in this region the dust loading results in heating of the system. In turn, the region below the curve is where dust results in system cooling. For instance, if the WMO model¹³ (curve 4) is used, heating occurs over surfaces with an albedo larger than 0.05 for cosine of solar zenith angle higher than 0.6. The calculations were performed by running a one-dimensional model using a two-stream radiation code with detailed wavelength resolution¹⁴.

needs to be determined before climate models can simulate the changing climate in the industrial era. □

Received 10 November 1995; accepted 30 April 1996.

- Houghton, J. T. et al. (eds) *Climate Change 1994: Radiative Forcing of Climate and an Evaluation of the IPCC IS92 Emission Scenarios* (Cambridge Univ. Press, 1994).
- Andreae, M. O. in *World Survey of Climatology* Vol. 16, *Future Climates of the World* (ed. Henderson-Sellers, A.) (Elsevier, Amsterdam, 1995).
- Tegen, I. & Fung, I. J. *geophys. Res.* **100**, 18707–18722 (1995).
- Tegen, I., Lacis, A. A. & Fung, I. *Nature* **380**, 419–422 (1996).
- Charlson, R. J. et al. *Tellus* **43A**, 152–158 (1991).
- Charlson, R. J. et al. *Science* **255**, 423–430 (1992).
- Penner, J. E., Dickinson, R. E. & O'Neill, C. A. *Science* **256**, 1432–1433 (1992).
- Chylek, P. & Wong, J. G. D. *Geophys. Res. Lett.* **22**, 929–931 (1995).
- Li, X., Maring, H., Savoie, D., Voss, K. & Prospero, J. M. *Nature* **380**, 416–418 (1996).
- Fouquart, Y., Bonnel, B., Chaoui Roquai, M. & Santer, R. J. *Clim. appl. Met.* **26**, 28–52 (1987).
- Carlson, T. N. & Benjamin, S. J. *atmos. Sci.* **37**, 193–213 (1980).
- D'Almeida, G. J. *geophys. Res.* **92**, 3017–3026 (1987).
- WMO Report of the Experts Meeting on Aerosols and their Climatic Effects (Rep. WCP-55, World Climate Program, Geneva, 1983).
- Sokolik, I. N. & Golitsyn, G. S. *Atmos. Envir.* **27A**, 2509–2517 (1993).
- Ackerman, S. A. & Cox, S. K. J. *geophys. Res.* **87**, 8991–9002 (1982).
- Sagan, C., Toon, O. B. & Pollack, J. B. *Science* **206**, 1356–1368 (1979).
- United Nations Environment Programme *World Atlas of Desertification* (Edward Arnold, London, 1992).
- Goudie, A. S. *Prog. phys. Geogr.* **7**, 502–530 (1983).
- Schutz, L. *Ann. N.Y. Acad. Sci.* **338**, 515–532 (1980).
- Duce, R. A. in *Aerosol Forcing of Climate* (eds Charlson, R. J. & Heintzenberg, J.) 43–72 (Wiley, New York, 1995).
- Andronova, A. V. et al. *Izv. Atmos. Oceanic Phys.* **25**, 29–36 (1989).
- Penner, J. E. et al. *Bull. Am. met. Soc.* **75**, 375–400 (1994).
- Ackerman, S. A. & Chung, H. J. *appl. Met.* **31**, 223–241 (1992).
- Kiehl, J. T. & Rodhe, H. in *Aerosol Forcing of Climate* (eds Charlson, R. J. & Heintzenberg, J.) 281–296 (Wiley, New York, 1995).
- Harrison, E. F. et al. *J. geophys. Res.* **95**, 18687–18698 (1990).
- Prospero, J. M., Uematsu, M. & Savoie, D. L. in *Chemical Oceanography* (eds Riley, J. P., Chester, R. & Duce, R. A.) 188–218 (Academic, London, 1989).
- Westphal, D. L., Toon, O. B. & Carlson, T. N. *J. atmos. Sci.* **45**, 2145–2166 (1988).
- Tegen, I. & Fung, I. J. *geophys. Res.* **99**, 22897–22914 (1994).
- Sokolik, I. N., Andronova, A. V. & Johnson, T. C. *Atmos. Envir.* **27A**, 2495–2502 (1993).
- Atmospheric Turbidity and Precipitation Chemistry Data (WMO, US Dept of Commerce and NOAA, Asheville, NC, 1972–75).
- Ben Mohamed, A. et al. *J. appl. Met.* **31**, 1286–1293 (1992).
- Panchenko, M. V. et al. *Atmos. Envir.* **27A**, 2503–2511 (1993).
- Kondratiev, K. Ya. & Zhvaley, V. F. *The First GARP Global Experiment: Aerosol and Climate* (in Russian) (Hydrometeoizdat, Leningrad, 1981).
- Stowe, L. L. in *Proc. AGU Spring Meeting* May 23–27, Baltimore, Maryland (Am. Geophys. Union, 1994).

ACKNOWLEDGEMENTS. I. Sokolik was a National Academy of Sciences Research Associate. This work was supported by NASA Radiation Sciences programme managed by R. Curran.

A large deep freshwater lake beneath the ice of central East Antarctica

A. P. Kapitsa*, J. K. Ridley†, G. de Q. Robin‡, M. J. Siegert§ & I. A. Zotikov||

* Faculty of Geography, Moscow State University, Moscow, Russia

† Mullard Space Science Laboratory, University College London, Dorking, Surrey, RH5 6NT, UK

‡ Scott Polar Research Institute, University of Cambridge, CB2 1ER, UK

§ Centre for Glaciology, Institute of Earth Studies, University of Wales, Aberystwyth, Dyfed SY23 3DB, UK

|| Institute of Geography, Russian Academy of Sciences, Moscow, Russia

IN 1974–75, an airborne radio-echo survey of ice depths over central East Antarctica led to the discovery of a sub-ice lake of unknown depth and composition, with an area of about 10,000 km² and lying beneath ~4 km of ice¹. In 1993, altimetric data from satellite measurements² provided independent evidence of the lake's areal extent, thus confirming it to be the largest known sub-ice lake by an order of magnitude. Here we analyse new altimetric and radio-echo data, along with existing seismic data³, to show that the lake is deep (mean depth of 125 m or more) and fresh, and that it has an area that exceeds previous estimates by about 50%—dimensions comparable with those of Lake Ontario. We estimate that the residence time of the water in the lake is of the order of tens of thousands of years, and that the mean age of water in the lake, since deposition as surface ice, is about one million years. Regional ice-dynamics can be explained in terms of steady-state ice flow along and over the lake.

The satellite ERS-1, launched in 1991, was the first to be specifically programmed for altimetric surveys of large polar ice sheets. An analysis of the initial fast delivery data gave mean surface elevations every 6.7 km along the satellite track². These defined areas of level ice, which roughly matched the lake boundary indicated by radio-echo sounding (RES), but differed by up to 30 km on the north-west of the lake and did not include Vostok, the Russian ice-drilling station (Fig. 1b).

The new lake boundary (Fig. 1b) is based on changes of surface slope derived from the waveform product⁴ using the latest altimeter data. This agrees with the lake boundary shown by RES profiling (Fig. 1a) to within 5 km (maximum error in aircraft position from the inertial navigation⁵).

Owing mainly to the high noise level, seismic shooting studies of the ice cover of central Antarctica produced no evidence, before 1993, of the meltwater lenses beneath the ice sheet that had been suggested by Kapitsa⁶ on morphological grounds. Although increasing shot depths^{7,8} and higher frequency filtering⁹ helped to reduce seismic noise levels, the only major reduction of noise on the Antarctic plateau had been achieved once by Kapitsa and Sorochin³ in 1964 by using a vertical seismometer spread from 2.5 to 49 m depth in a borehole at Vostok (Fig. 2). This clearly confirmed the ice depth as 3,700 m for the important ice-coring programme at Vostok, whereas a later echo was then interpreted as a sedimentary layer.

Confirmation of the ice depth and the presence of a water layer by RES profiling (Fig. 3a) led to reinterpretation of the seismic data shown in Fig. 2. The absence of any significant energy return between reflections from the ice–water interface and the lake floor on seismometers below 10 m depth, identification of the later echoes as P waves (Fig. 2), together with evidence from RES and altimetry, all confirm the presence of a deep water layer. Assuming 1,450 ms⁻¹ as the P-wave velocity in water gives the water thickness as 510 m and the lake bed at 710 m below sea level.

Along RES flight lines, the depth of floating ice changes slowly. The lake bed is around 700 m below sea level near both ends of the lake. Assuming this applies over the whole lake bed, the mean

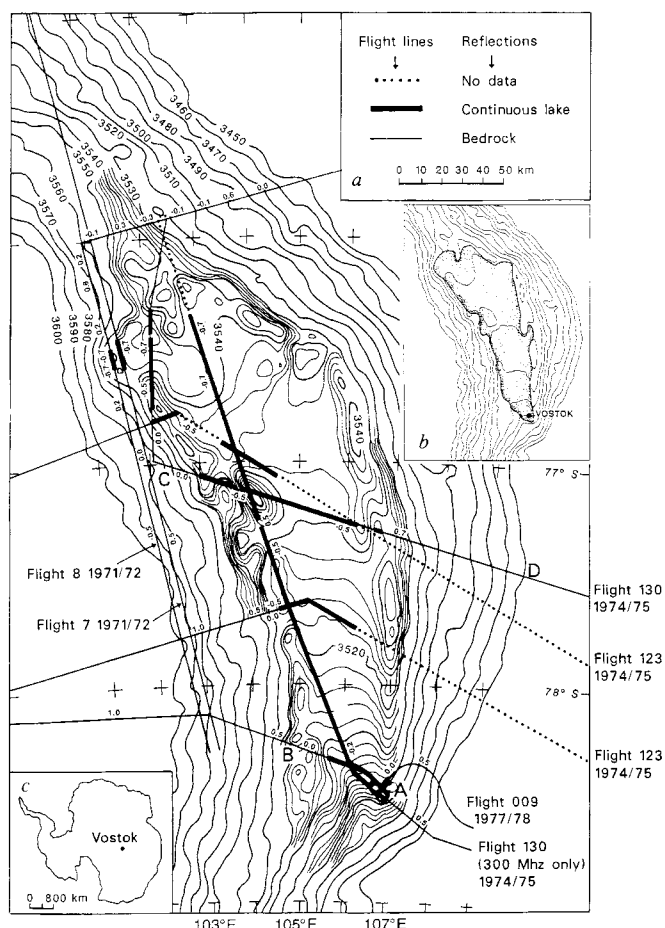


FIG. 1 Surface topography. ERS-1 waveform product (WAP) altimeter maps of the Vostok region. The improved altimetry comes from two 168-day repeat cycles of satellite orbits. The WAP analysis⁴ gave 600,000 mean surface elevations over circles of 3 km diameter. Repeated and overlapping measurements gave an r.m.s. deviation of 20 cm and a maximum relative error in contours of 50 cm for the first time. This can now be extended over all inland ice north of 82° S. a, Contours over and near the lake boundary at 2-m intervals, with 10-m contours over steeper slopes. Figures along flightlines show bedrock elevations in kilometres above or below (—) sea level. Unpublished data now included come from flight 009 of 1977/78 and spot soundings every 2.4 km along the south-to-north flight 130 of 1974/75, which was not covered by continuous profiling (Fig. 3a). b, Shaded area shows the extent of the sub-ice lake, based on 2-m contouring but with 10-m contours over the entire area. This shows the low surface slope over the lake compared with the surrounding surface topography. c, Location of Vostok station.

water thickness beneath the south–north profile of flight 130 is 125 m and the total volume of water 1,800 km³. The bedrock topography around the lake is similar to the topography around the rift valleys, especially of the deglaciated Lake Baikal, Siberia. Similar lake bathymetry would considerably increase the estimated volume of the sub-ice lake.

With a surface accumulation of 2.7 g cm² yr⁻¹, an ice thickness of ~3,700 m and a geothermal flux of ~50 mW, steady-state theory^{9,10} predicts basal melting of ~1 mm yr⁻¹. Glacier dynamic theory indicates that the time taken for ice deposited on the surface to melt at the base is ~10⁶ years. These estimates are in broad agreement with the age of the ice core at Vostok¹¹, shown by isotope profiling to a depth of 2,700 m.

Under steady-state conditions, a water input of 1 mm yr⁻¹ from floating ice and a water layer 125 m thick gives a residence time of 1.25 × 10⁵ years. Increasing the water input from basal melting of surrounding grounded ice and increasing the volume of the water in the same proportion leaves the residence time unchanged. With

Two-way travel time (s)

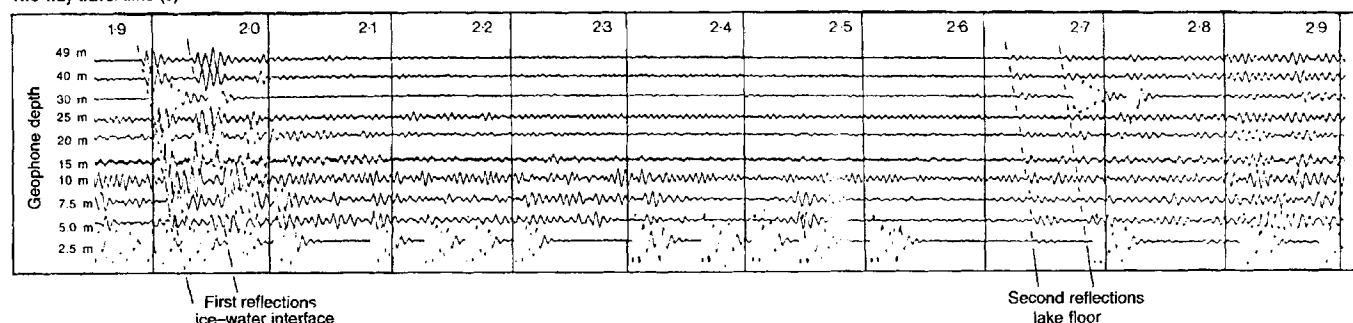


FIG. 2 Section of 24-channel seismic record from Vostok relevant to the sub-ice water layer. This records movement over a vertical line of seismometers from 49 m to 2.5 m depth in a borehole 180 m from the explosion. It covers a period from 1.85 to 2.9 seconds after the explosion of 5 kg of TNT at 39 m depth and 180 m from the vertical seismometer line. A conventional horizontal spread of twelve seismometers at 20-m intervals recorded the same echoes (not shown) at ~ 1.92 and ~ 2.65 s against a much higher background noise. The echo from the base of the floating ice reaches the deepest (49 m) seismometer first at ~ 1.91 s. It then travels up the seismometer line to the surface where it is reflected down to pass the

49-m seismometer ~ 50 ms later. This has a mean velocity of $\sim 2,200$ ms $^{-1}$, typical of compression (P) waves in the top 50 m of firn in this region¹². About 45 ms after the first arrival, a second wave train of similar intensity and duration follows as a result of the initial surface reflection of the explosion at 39 m depth. There is no significant return of energy between ~ 2.00 and ~ 2.63 s, when a weaker wave train passes up and down the seismometer line at the same velocity. This confirms that they are compressive (P) waves, the only waves that travel through water, and not transverse (PS) waves, which are sometimes recorded from shots on ice shelves.

both likely, our best estimate of the residence time of the lake water is of the order of 50,000 years.

Other features of the water mass can be deduced from the relationship between surface elevation and thickness of floating ice (Fig. 4). The close scatter of points around the freshwater line (Fig. 4) indicates a hydraulic pressure equivalent to a head of water close to 3,140 m above sea level over the whole lake. It also confirms the presence of relatively fresh water in the lake. Any difference between the mean slope of plotted points and the freshwater line could be due to residual effects from boundary stresses, to errors of ice thickness and/or to the effect of limited salinity on water density within a range of 0.00 and 0.05‰.

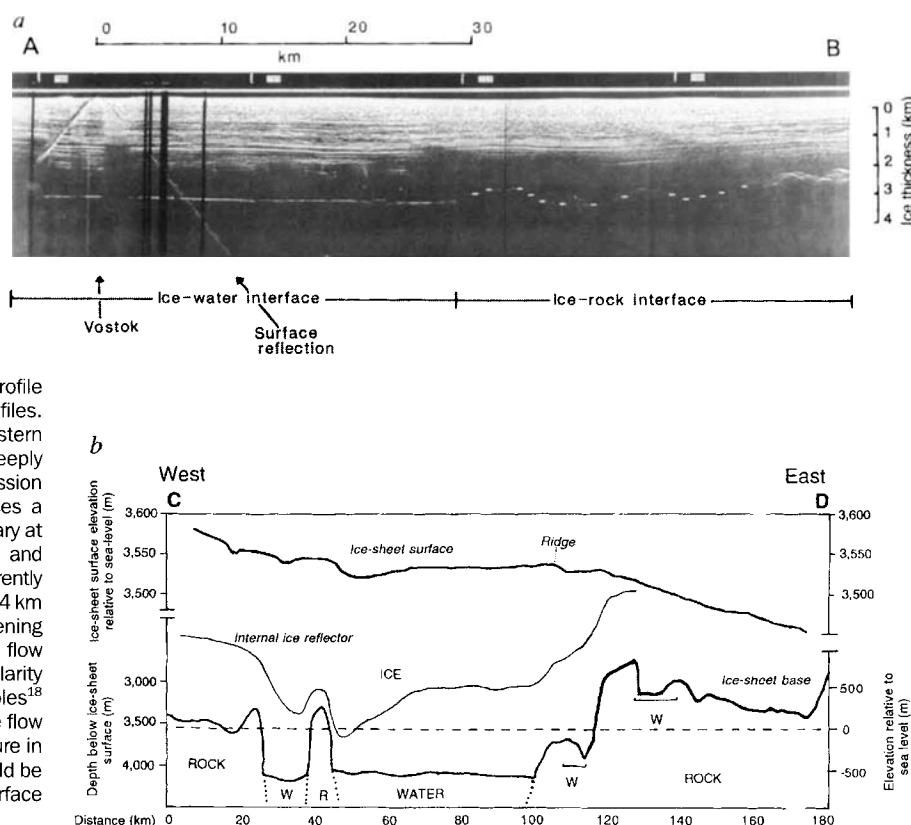
The decrease of ice thickness of ~ 500 m from north to south of

the lake means that the ice–water contact temperature will be 0.4°C lower at the north than at the southern end. This may be sufficient to drive slow circulation of sub-ice water and affect basal heat flow distribution over the lake, but not the total heat flux.

Ice cores from Vostok contains micro-organisms^{12,13} carried by air to Antarctica on dust particles from low latitudes. Melting of basal ice will release these microorganisms into lake water around one million years after deposition on the ice surface. Sediments on the lake bed may be several million years older, an exceptional environment that should provide useful information to the biological community and to geologists.

The lake occupies the lower part of a basin or rift (Figs 1, 3b). Unlike grounded ice, basal drag is negligible over an ice–water

FIG. 3 Two-dimensional vertical profiles. *a*, Continuous RES profile from flight 009, 1977–78, along line A–B in Fig. 1a. Strong ice–water reflections beneath Vostok and for ~ 30 km give way to weaker echoes from ice–bedrock further west. White dots have been added to show the location of weak bedrock echoes where they are not clearly visible. The near-horizontal layering in the upper 2 km is due to higher-conductivity layers of ice resulting from widespread deposition of volcanic material over the ice sheet¹⁷. Slanting echoes show reflections from surface buildings as the aircraft passed over Vostok. *b*, Cross-section along line C–D in Fig. 1a, showing deepest continuous RES layer. The vertical scale of the surface profile is increased to 100 times that of the deep profiles. Surface valleys around 30 and 50 km near the western boundary are due to downslope motion over steeply falling bedrock, opposed by horizontal compression that continues over the lake. After the ice crosses a rock ridge approximately parallel to the lake boundary at 40 km, a deeper valley is formed. The surface and reflecting layer then rise from 50 to 70 km, apparently above faster, southward-moving ice. From 70 to 94 km the surface is almost level, with slight ice thickening probable, until the lake boundary is crossed. A flow component to the east is then indicated by the similarity of the profile in *b* to that across the Doake Ice Rumples¹⁸ (77.7°S , 66.6°W) to the Ronne Ice Shelf. Upslope flow appears to be initiated and driven by vertical pressure in both cases. The alternative to this flow pattern would be an ice divide, with static ice below an almost flat surface for at least 10 km to either side.



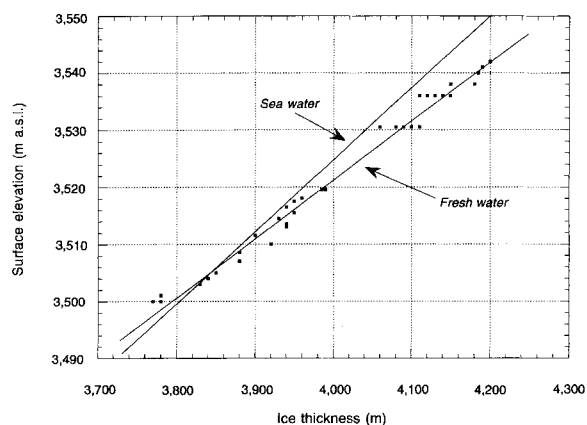


Fig. 4 Thickness (h) of floating ice by RES plotted against surface elevation (E) at locations clear of boundary effects along and across the lake on flight 130. Elevation of corresponding points is interpolated from satellite altimetry (Fig. 1). Also shown are computed slopes for ice in the same thickness range given by $dE/dh = (1 - \rho_{ice}/\rho_{water})$, where ρ denotes density. The lines shown apply to pressures from 30 to 40 MPa (ref. 19), salinity of sea water²⁰ of 35‰, and a mean density of ice of 0.913 g cm^{-3} from the borehole at VS (J.-R. Petit, personal communication). Resultant slopes are 0.105 for fresh water and 0.127 for sea water. Relative errors in ice thickness measurements should not exceed 30 m.

interface. Away from boundary effects, ice is compressed vertically and expands horizontally in directions offering least resistance. This produces a low surface slope normal to the surface contours as on ice shelves^{14,15}. At Vostok¹⁶, flow is normal to local contours. In general, contours show flow along the length of the lake, starting eastwards at 76.5°S and swinging southwards by 77.0°S . The near-parallel tilt of the deep reflecting layer to the valley walls (Fig. 3b) is also typical of flow in the direction of a long valley¹⁷.

The above flow requires a high influx of ice over the northwest boundary. Here steep surface slopes, thick grounded ice with basal melting and sliding, and ice-sheet contours to the west, all indicate a large inflow at around 76.5°S .

Along almost all the eastern boundary, the ice surface rises $\sim 8 \text{ m}$ in 10 km before falling at the regional slope to the east, still over rising bedrock (Fig. 3b). This is similar to profiles across the Ronne Ice Shelf rumples¹⁸. In both cases, the initial rise extends over three times the ice thickness (h). Although magnitudes of ice thickness, slopes and velocities differ considerably, the same basic mechanism driven by vertical compression seems to apply.

The loss of ice over the eastern boundary will decrease velocity towards the south. To maintain vertical compression at around a constant level in order to drive flow to the south and east requires an increased surface slope in proportion to the slowing of velocity. Narrowing of the lake southwards, as with ice shelves, also requires increased surface slope.

Estimates of the easterly ice movement onto the lake, based on steady-state discharge of accumulation from the ice divide (dome B) 240 km to the west, give a mean inflow at $\sim 3 \text{ m yr}^{-1}$. The only direct measurement of ice velocity in the region was made at Vostok¹⁶ by star sights in 1964 and 1972. This gave $3.7 \pm 0.7 \text{ m yr}^{-1}$ at $142 \pm 10^\circ$ true, approximately normal to contours within 20 km of the station. This gives an eastward component of $2.3 \pm 0.5 \text{ m yr}^{-1}$. When the loss of ice across the eastern boundary is considered, the flow at Vostok equals that expected for steady-state flow to well within the error of estimates and measurements.

Further geophysical studies of the lake are being planned. Satellite global positioning system measurements of surface velocity in three dimensions will show how closely steady-state flow applies from changes of surface elevation with time, while measured horizontal motion will test predictions of flow from ice dynamics. Seismic sounding and RES survey will give a much

better knowledge of the lake depth and extent and determine the nature of any lake bed sediments. Finally, sampling of lake water and sediments should be undertaken, subject to Environmental Impact Assessment under the Antarctic Treaty, when this can be achieved without polluting the lake water. □

Received 2 September 1995; accepted 16 May 1996.

1. Robin, G. de Q. et al. *Phil. Trans. R. Soc. Lond. B* **179**, 185–196 (1977).
2. Ridley, J. K., Cudlip, W. & Laxon, S. W. *J. Glaciol.* **39**, 625–634 (1993).
3. Kapitsa, A. P. & Sorochtin, O. G. *Sov. Antarct. Eksped. Info. Bull.* **51**, 299–302 (1965). (English transl.).
4. ERS User Handbook, SP-1148 Revision 1 (ESA, Noordwijk, The Netherlands, 1993).
5. Rose, K. E. thesis, Univ. Cambridge, 1978).
6. Kapitsa, A. P. *Trudy Sov. Antarct. Eksped.* **18**, 36–37 (1961). (in Russian).
7. Robin, G. de Q. *Norw.-Br.-Swed. Antarct. Exped.* **VIII** (Norsk Polar Institute, Oslo, 1958).
8. Kapitsa, A. P. & Sorochtin, O. G. *Int. Ass. Sci. Hydrol. publ. no. 61*, 162–164 (1963).
9. Robin, G. de Q. *J. Glaciol.* **2**, 523–532 (1955).
10. Zotikov, I. A. Academy of Sciences, USSR, Interdepartmental Commission on Antarctic Research, Commission Reports, 22–36 (1961).
11. Lorius, C. et al. *Nature* **316**, 591–596 (1985).
12. Oswald, G. K. A. & Robin, G. de Q. *Nature* **245**, 251–254 (1973).
13. Abyzov, S. S. *Antarctic Microbiology* (ed. Freidmann, E. I.) 265–295 (Wiley-Liss, New York, 1993).
14. Paterson, W. S. B. *The Physics of Glaciers* 3rd edn (Pergamon, Oxford, 1994).
15. Robin, G. de Q. *Nature* **253**, 168–172 (1975).
16. Liebert, J. *Cal. R. Oss. astr. Coll. romano Ramal (suppl.)* **10**, 381–382 (Vermessungstechnik, Berlin (1973)).
17. Robin, G. de Q. & Millar, D. H. M. *Ann. Glaciol.* **3**, 290–294 (1982).
18. Smith, A. M. *Br. Antarct. Surv. Bull.* **72**, 47–52 (1986).
19. Dorsey, N. E. *Properties of Ordinary Water Substances* (Am. Chem. Soc. Monogr. ser. no. 81, Reinhold, New York, 1940).
20. Sverdrup, H. U., Johnson, M. W. & Fleming, R. H. *The Oceans* (Prentice Hall, New York, 1942).

ACKNOWLEDGEMENTS. Our observations were made possible by massive logistic support of Antarctic operations by air, land, sea and satellite. We thank all Antarctic organizations in the United States, Russia and Europe, as well as our home institutions and SCAR for all their extensive and efficient support.

CORRESPONDENCE and requests for materials should be addressed to M.J.S. (e-mail: mas@aber.ac.uk).

Geochemistry of mantle–core differentiation at high pressure

Jie Li & Carl B. Agee

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

THE apparent excess of siderophile (iron-loving) elements in the Earth's mantle has been a long-standing enigma in the geochemistry of mantle–core differentiation^{1,2}. Although current models have proved successful in explaining some aspects of this problem^{3–7}, important questions remain. In particular, the mantle's near-chondritic ratio of nickel to cobalt (close to that expected for the material from which the Earth formed) is hard to explain, given the markedly different ambient-pressure partitioning behaviour of these elements between iron-alloy and silicate melts^{3–8}. Here we report experimental results which show that both elements become less siderophile with pressure, but the effect is much more pronounced for Ni, so that the partition coefficients of the two elements become essentially equivalent at an extrapolated pressure of $\sim 28 \text{ GPa}$. The absolute and relative abundances of Ni and Co in the mantle are therefore consistent with alloy–silicate chemical equilibrium at high pressure, indicating that core formation may have taken place in a magma ocean with a depth of $750\text{--}1,100 \text{ km}$. We also find that, unlike Ni and Co, sulphur becomes more siderophile with pressure. Sulphur's increased affinity for iron with depth could make it the dominant light element in the Earth's core.

Early studies on core–mantle formation based on low-pressure partitioning data suggested an apparent disequilibrium between the mantle and the core^{1,2,5,6,8}. Moderately and highly siderophile elements are far in excess in the mantle with respect to what would be expected from these data. Moreover some elements with very different low-pressure partition coefficients have near-chondritic

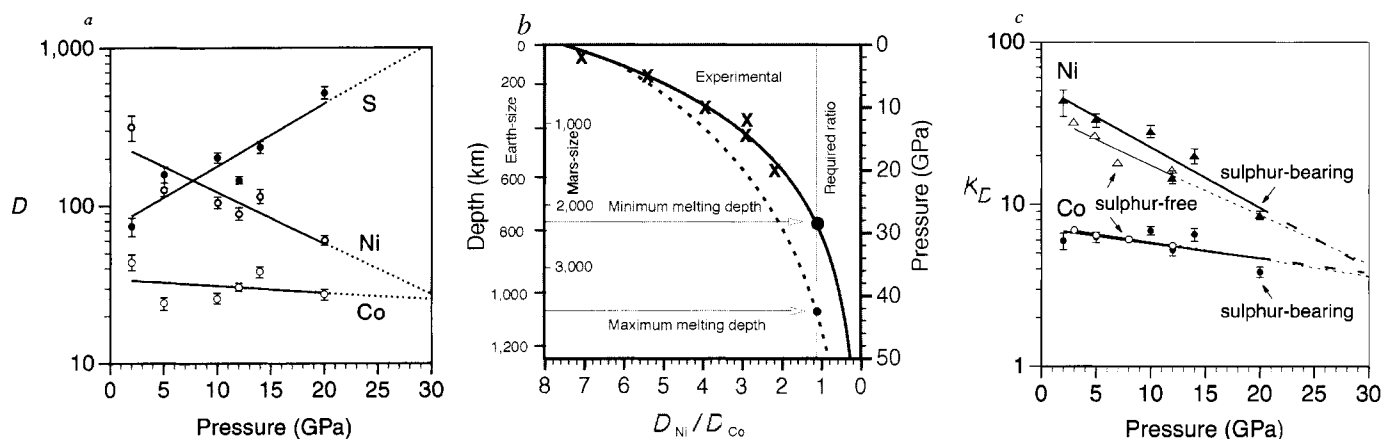


FIG. 1a, Partition coefficient D (alloy liquid/silicate liquid) versus pressure diagram for Ni, Co and S. Lines are least-squares fits of the form $\log D = aP + b$, where a and b are constants. Error bars are calculated from the error of the mean of the analyses. b, $D_{\text{Ni}}/D_{\text{Co}}$ versus pressure diagram. The equivalent depths in the Earth and Mars are shown as references. Heavy solid line is a least-squares fit to the experimental data (crosses) of the form $D_{\text{Ni}}/D_{\text{Co}} = a \exp(bP)$, where a and b are constants. Note that the fit is better than in a as the effect of variation in f_{O_2} cancels out in $D_{\text{Ni}}/D_{\text{Co}}$. This solid line describes the changing $D_{\text{Ni}}/D_{\text{Co}}$ with depth in a magma ocean. It intersects the required-ratio line ($D_{\text{Ni}}/D_{\text{Co}} = 1.1$) at ~ 28 GPa. Dotted line is calculated from experimental data and represents the bulk effect of local

equilibrium between a compositionally zone d magma ocean and the core. It intersects the required ratio line at ~ 42 GPa. c, Exchange coefficient $K_D^{\text{Ni,Co}}$ versus pressure diagram comparing our data with that of Thibault and Walter¹² who studied a sulphur-free system. All lines are least-squares fits of the form $\log K_D^{\text{Ni,Co}} = aP + b$, where a and b are constants. Note that the fits are better than those in a after correcting for f_{O_2} variations. Solid lines represent best-fits experimental results. Dashed and dotted lines are extrapolations to higher pressure for sulphur-bearing and sulphur-free data, respectively. The highest-pressure experiment of Thibault and Walter falls within the error bars of our value for $K_D^{\text{Ni}} = 15 \pm 1$ and $K_D^{\text{Co}} = 5.2 \pm 0.4$ at 12 GPa. Error bars are calculated as in a.

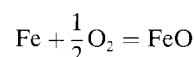
ratios, for example Ni/Co and Re/Ir (ref. 2). Although low-pressure models can account for the excess and the near-chondritic ratio of highly siderophile elements such as Re and Ir, they do not give a self-consistent explanation for the near-chondritic ratio of moderately siderophile elements such as Ni and Co (refs 3, 5–7).

It has been proposed that element partitioning during mantle–core segregation was likely to have operated at very high temperatures and high pressures, probably during the deep magma ocean stage of early Earth differentiation when molten alloy and molten silicate underwent gravitational separation^{4,7}. Several experimental studies^{10–13} have concluded that a simple temperature effect probably cannot explain the distribution of siderophile elements between the Earth's mantle and core. For example, isobaric experiments between 2,000 and $\sim 2,500^\circ\text{C}$ have shown a negligible temperature effect on partitioning of Ni and Co between molten alloy and silicate^{12,13}. On the other hand, the same studies indicate that pressure may partially account for the excess siderophiles in the mantle. The data set presented in this Letter show that the pressure effect is actually a key ingredient in resolving the mantle siderophile element problem.

Experiments were performed at pressures of 2–20 GPa, under isothermal conditions of $2,000^\circ\text{C}$, with a Walker-style octahedral multi-anvil device. Details of the experimental design and technique can be found in Agee *et al.*¹³. Starting material was finely ground powder from a split of the Allende meteorite¹³, but the bulk composition of the run products were slightly different from that of Allende owing to exchange with the MgO capsules. All experiments contained coexisting immiscible liquids of silicate and Fe-rich alloy. Both silicate and alloy liquids, on quenching, formed discrete, relatively large masses of crystals and glass. The average composition of these domains were determined by multiple broad beam analyses using a Cameca electron microprobe. A summary of the data and experimental conditions are given in Table 1. Element partitioning between molten alloy and silicate is presented in the form of the partition coefficient D_i ($D_i = C_i^{\text{AL}}/C_i^{\text{SL}}$, where C_i^{AL} and C_i^{SL} are the concentrations of element i in weight per cent in the alloy liquid and the silicate liquid, respectively). The results are illustrated in Fig. 1a. In the range 2–20 GPa D_{Ni} decreases more than five times, from 318 to

59. Over the same range D_{Co} only decreases by less than half from 45 to 27. We interpret these trends as evidence that Ni and Co become less siderophile with pressure in the deep mantle. Furthermore, the trend of D_{S} in Fig. 1a argues strongly that sulphur becomes more siderophile with pressure.

It is well known that siderophile element partitioning between Fe-alloys and silicates is sensitive to oxygen fugacity (f_{O_2} ; refs 14–17). Theoretically f_{O_2} , relative to iron-wüstite oxygen fugacity buffer (IW), can be estimated based on Fe metal content of the alloy liquid and FeO content of the silicate liquids, respectively¹⁴:



$$\log f_{\text{O}_2} = \log f_{\text{O}_2}(\text{IW}) + 2 \log a_{\text{FeO}}^{\text{SL}} - 2 \log a_{\text{Fe}}^{\text{AL}}$$

where $a_{\text{FeO}}^{\text{SL}}$ is the activity of FeO in silicate liquid, and $a_{\text{Fe}}^{\text{AL}}$ is the

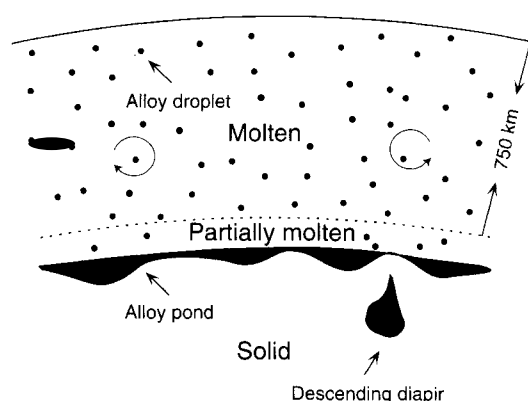


FIG. 2 Cross-section cartoon (inspired by Stevenson²⁵) of the 750-km-depth model illustrating a high-pressure superliquidus mantle–core formation scenario constrained by the Ni/Co data. Small alloy droplets settle rapidly through the strongly convective silicate liquid and enter an alloy pond near the partially molten layer at the bottom of the magma ocean. The composition of the mantle reflects the equilibration between alloy and silicate near the ponded alloy, before it descends into the growing core.

activity of Fe metal in alloy liquid. Following the method described in Table 1, we calculated the f_{O_2} in our experiments to be near of slightly below IW and varying only by a half log unit. This range of f_{O_2} is broadly consistent with the constraints on oxygen fugacity relevant to mantle–core formation^{4,16–18}. But the assumptions used in the calculations remain to be tested, hence the true f_{O_2} might be different from the calculated value. Fortunately, if Fe, Ni and Co all exist as divalent cations in silicate melt and behave ideally, the effect of f_{O_2} on D_{Fe} , D_{Ni} and D_{Co} should largely cancel out in the ratio D_{Ni}/D_{Co} and the exchange coefficient $K_D^{Ni,Co}$ ($K_D^{Ni,Co} = D_{Ni,Co}/D_{Fe}$; refs 19, 20) (refs 12, 16, 17). Consequently D_{Ni}/D_{Co} and $K_D^{Ni,Co}$ are relatively insensitive to f_{O_2} and the observed variations in D_{Ni}/D_{Co} and $K_D^{Ni,Co}$ are owing mainly to a pressure effect that holds for a wide range of f_{O_2} (Fig. 1b, c).

Another factor to consider is the effect of sulphur on element partitioning^{19,21}. To evaluate the importance of this effect, we compare our data with that of Thibault and Walter¹² who studied partitioning of Ni and Co between molten Fe-alloy and silicate melt in a sulphur-free system with graphite capsules at 2,227 °C. Figure 1c shows that there is little difference in K_D^{Ni} and K_D^{Co} between sulphur-free and sulphur-bearing systems at high pressure. Therefore, we feel confident that our data are applicable to high-pressure, superliquidus mantle–core formation scenarios with variable sulphur content.

There are a number of ways that the high-pressure convergence of D_{Ni} and D_{Co} can be applied to mantle–core formation scenarios. Mass-balance calculations²² have demonstrated that the upper mantle's near-chondritic Ni to Co ratio (Ni/Co) can be produced from a bulk Earth with chondritic Ni/Co if D_{Ni} and D_{Co} are ~27

and ~24 respectively ($D_{Ni}/D_{Co} = 1.1$). Our experiments show that D_{Ni}/D_{Co} reaches the required value of 1.1 at a pressure of ~28 GPa (Fig. 1b). D_{Ni} and D_{Co} are 29 and 26 respectively at this pressure (Fig. 1a), remarkably close to the absolute D -values governed by mass balance. The pressure of 28 GPa can be interpreted as the base of the molten mantle (depth ~750 km, temperature ~2,100 to ~2,400 °C; refs 13,23) where a well-mixed magma ocean equilibrates with ponded molten alloy, if such a deep magma ocean ever existed (Fig. 2). How does the ponded Fe-alloy reach the core? It must sink through the solid deeper mantle. If the ponded alloy sinks rapidly in large diapirs, then chemical re-equilibration with the deeper mantle may be negligible. If sinking is by percolation through solid silicate matrix, then the core alloy may re-equilibrate with the deeper mantle on the way down.

Another way of interpreting the D_{Ni}/D_{Co} convergence to 1.1 is by considering the average or bulk value of D_{Ni}/D_{Co} for an entire magma ocean column, from surface to bottom. In this case the magma ocean would be deeper, and the bottom pressure would be greater than 28 GPa. This comes from averaging low-pressure D_{Ni}/D_{Co} ratios (that are greater than 1.1) with high-pressure D_{Ni}/D_{Co} (that are less than 1.1). Constraining this bulk D_{Ni}/D_{Co} to equal 1.1 requires a magma ocean with a maximum pressure of ~42 GPa or a depth of ~1,100 km (Fig. 1b). The physical meaning of this 'bulk value depth' could be small immiscible alloy droplets dispersed throughout a poorly mixed magma ocean. In contrast to the 750-km-depth scenario, here convective mixing is inefficient—the silicate melt column equilibrates with droplets at local pressures as they sink and develops a Ni/Co gradient that increases with depth. Ultimately over geological time, the gradient must be

TABLE 1 Experimental conditions, electron microprobe analyses and results

Run no.	516A8	477A8	472A8	508A6	482A6	515A6
Pressure (GPa)	2	5	10	12	14	20
Temperature (°C)	2,000	2,000	2,000	2,000	2,000	2,000
Duration (min)	2.5	3	3	5	5	15
Composition of silicate liquid (wt%)						
Number of analyses*	23(6)	± [†] 16(8)	± [†] 13(10)	± [†] 10(5)	± [†] 10(6)	± [†] 28(7)
Si	17.32	0.12	16.30	0.13	16.47	0.07
Mg	27.69	0.27	22.89	0.38	20.65	0.19
Al	1.46	0.10	1.98	0.10	2.42	0.05
Ca	2.41	0.16	2.27	0.15	3.27	0.07
Fe	7.22	0.26	14.88	0.41	15.38	0.11
Ni	0.047	0.008	0.11	0.01	0.13	0.01
Co	0.010	0.001	0.019	0.001	0.018	0.001
S	0.39	0.05	0.17	0.02	0.14	0.01
Composition of alloy liquid (wt%)						
Number of analyses	7	10	13	12	12	14
Fe	53.58	0.90	57.46	0.34	57.33	0.35
Ni	14.85	0.86	14.00	0.25	13.65	0.44
Co	0.44	0.02	0.46	0.02	0.46	0.01
S	28.76	0.61	26.96	0.43	28.32	0.36
D (alloy/silicate)						
Fe	7.4	0.3	3.9	0.1	3.7	0.0
Ni	316	57	127	12	105	9
Co	44	5	24	2	26	2
S	74	10	159	19	202	15
f_{O_2} (in log unit relative to IW)‡	-0.5	-0.1	0.0	-0.5	-0.4	-0.4

For major elements, a beam current of 30 nA and a counting time of 10–20 s on peaks and backgrounds were used. For minor elements, a beam current of 30 nA and a counting time of up to 100 s on peaks and backgrounds were used. Special care was taken for Co, which is a trace element in the silicate liquid. Several standards including an olivine standard with 138 p.p.m. of Co (ref. 29) were used for calibration. Co in the silicate liquid was measured simultaneously on two spectrometers with a beam current of 140 nA and a counting time of 300 s on peaks and backgrounds.

* Values in parentheses are the number of analyses for Co.

† Standard error of the mean.

‡ In calculating f_{O_2} , we assume (1) ideal mixing so that $a_{FeO}^{SL} = X_{FeO}^{SL}$ and $a_{Fe}^{AL} = X_{Fe}^{AL}$, where X_{FeO}^{SL} is the mole fraction of FeO in silicate liquid and X_{Fe}^{AL} is the mole fraction of Fe metal in alloy liquid; (2) iron and nickel in the molten alloy are present as monosulphide or metal³⁰; (3) small fractionations between Ni and Fe can be ignored¹⁹ so that $X_{Fe}^{AL} = X_{Fe}^{AL} - X_S^{AL} \cdot X_{Fe}^{AL} / (X_{Fe}^{AL} + X_{Ni}^{AL})$ where X_{Fe}^{AL} , X_S^{AL} and X_{Ni}^{AL} are the mole fractions of total iron, sulphur and nickel in the alloy liquid respectively.

erased by convective homogenization to produce a uniform, near-chondritic Ni/Co in the upper mantle.

We further consider a possible scenario where the whole mantle acquires a near-chondritic Ni/Co through high-pressure alloy/silicate equilibrium. Modern accretion theories suggest that core formation could have commenced when the Earth reached one-tenth of its present mass or about half of its present radius (~3,300 km, 'Mars-size')^{24–26}. In a proto-Earth of that size, the pressure at core–mantle boundary (CMB) would be ~28 GPa, or the pressure where $D_{\text{Ni}}/D_{\text{Co}} = 1.1$. As the Earth grew by accretion, the base of the magma ocean could have stayed near the 28-GPa isobar by migrating upwards. In other words, the outer layer of the mantle remained molten, but the deepest zone began to solidify up from the CMB. As a result, the whole mantle could have, at different times, equilibrated with the alloy at ~28 GPa. In this case, the upper and the lower mantles would have a similar Ni/Co signature.

It is interesting to note that our data suggest a magma ocean that extends just into the top of the Earth's lower mantle. This is consistent with the 'minimum melting depth' mass-balance calculation for refractory lithophile elements by Agee and Walker²⁷. The data do not favour a magma ocean that reached a depth of ~2,900 km (the present CMB) because mantle Ni/Co would become strongly super-chondritic. Shallow magma oceans (<700 km) are also ruled out because of the sub-chondritic Ni/Co produced at lower pressures.

We now consider the implications of sulphur becoming more siderophile with increasing pressure. Because of its moderately volatile nature, the fate of sulphur during accretion is more difficult to assess than elements such as Ni and Co. It is generally agreed that the sulphur content of the upper mantle is two orders of magnitude lower than the concentration of sulphur in chondrites^{2,28}. Such a large depletion can be partly explained by loss of volatiles to space during a high-temperature stage of Earth accretion. Alternatively, a significant fraction of the chondritic sulphur may reside in the Fe-alloy that makes up the Earth's core. Our experimental results support the notion of a sulphur-depleted mantle produced by alloy/silicate equilibria, and in turn support the hypothesis that sulphur is the dominant light element in the molten outer core. □

Received 13 December 1995; accepted 9 May 1996.

1. Ringwood, A. E. *Geochim. cosmochim. Acta* **30**, 41–104 (1966).
2. Jagoutz, E. et al. *Proc. Lunar planet. Sci. Conf.* **10**, 2031–2050 (1979).
3. Wänke, H. *Phil. Trans. R. Soc. Lond. A* **303**, 287 (1981).
4. Brett, R. *Geochim. cosmochim. Acta* **48**, 1183–1188 (1984).
5. Jones, J. H. & Drake, M. J. *Nature* **322**, 221–228 (1986).
6. Newsom, H. E. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 273–288 (Oxford Univ. Press, New York, 1989).
7. Murthy, R. V. *Science* **253**, 303–306 (1991).
8. Schmitt, W., Palme, H. & Wänke, H. *Geochim. cosmochim. Acta* **53**, 173–185 (1989).
9. Brett, R. *Geochim. cosmochim. Acta* **35**, 203–221 (1971).
10. Walker, D., Norby, L. & Jones, J. H. *Science* **262**, 1858–1860 (1994).
11. Hillgren, V. J., Drake, M. J. & Rubie, D. C. *Science* **264**, 1442–1445 (1994).
12. Thibault, Y. & Walter, M. J. *Geochim. cosmochim. Acta* **59**, 991–1002 (1995).
13. Agee, C. B. et al. *J. geophys. Res.* **100**, 17725–17740 (1995).
14. Drake, M. J., Newsom, H. E. & Capobianco, C. J. *Geochim. cosmochim. Acta* **53**, 2101–2111 (1989).
15. Capobianco, C. J. & Amelin, A. A. *Geochim. cosmochim. Acta* **58**, 125–140 (1994).
16. Dingwell, D. B., O'Neill, H. St C., Ertel, W. & Spettel, B. *Geochim. cosmochim. Acta* **58**, 1967–1974 (1994).
17. Holzheid, J. R., Borisov, A. & Palme, H. *Geochim. cosmochim. Acta* **58**, 1975–1981 (1994).
18. Arculus, R. A. *Rev. Earth planet. Sci.* **13**, 75–95 (1985).
19. Peach, C. L. & Mathez, E. A. *Geochim. cosmochim. Acta* **57**, 3013–3031 (1993).
20. O'Neill, H. St C. *Science* **257**, 1281–1285 (1992).
21. Jana, D. & Walker, D. V. M. *Goldschmidt Conf. Program and Abstracts* 58 (Geochemical Soc., 1995).
22. McDonough, M. F. & Sun, S.-s. *Chem. Geol.* **120**, 223–253 (1995).
23. Zhang, J. Z. & Herzberg, C. J. *geophys. Res.* **99**, 17729–17742 (1994).
24. Kaula, W. M. *J. geophys. Res.* **84**, 999–1008 (1979).
25. Stevenson, D. J. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 231–249 (Oxford Univ. Press, New York, 1989).
26. Tonks, W. B. & Melosh, H. J. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 151–174 (Oxford Univ. Press, New York, 1989).
27. Agee, C. B. & Walker, D. *Earth planet. Sci. Lett.* **90**, 144–156 (1988).
28. Lorand, J. P. *Geochim. cosmochim. Acta* **54**, 1487–1492 (1990).
29. Stosch, H.-G. *Contr. Miner. Petrol.* **78**, 166–174 (1981).
30. Sharma, R. C. & Chang, Y. A. *Metall. Trans.* **B10**, 103–108 (1979).

ACKNOWLEDGMENTS. We thank M. Walter and K. Righter for helpful comments. We also thank D. Lange for technical assistance in electron microprobe analyses. This work was supported in part by the US NSF.

The earliest known tyrannosaur from the Lower Cretaceous of Thailand

Eric Buffetaut*, Varavudh Suteethorn† & Haiyan Tong*

* Centre National de la Recherche Scientifique, URA 1761, Laboratoire de Paléontologie des Vertébrés, Université Paris 6, 4 place Jussieu, 75252 Paris Cedex 05, France

† Geological Survey Division, Department of Mineral Resources, Rama VI Road, Bangkok 10400, Thailand

THE TYRANNOSAURIDAE were the dominant large carnivorous dinosaurs in Asia (excluding India) and western North America during the Late Cretaceous period^{1–3}. Most of them are from the Campanian and Maastrichtian ages, and very little is known about their earlier history, although scanty remains have been reported from the early part of the Upper Cretaceous^{4–7}. We describe a newly discovered incomplete skeleton of a large theropod from the Early Cretaceous Sao Khua Formation of northeastern Thailand as an early and primitive representative of the Tyrannosauridae. This new taxon, which is at least 20 million years older than the earliest previously known tyrannosaurids, suggests that the early evolution of tyrannosaurids may have taken place in Asia.

Superorder Dinosauria Owen, 1842

Order Saurischia Seeley, 1887

Suborder Theropoda Marsh, 1881

Family Tyrannosauridae Osborn, 1905

Siamotyrannus isanensis gen. et sp. nov.

Etymology. Generic name from *Siam*, the old name of Thailand, and *tyrannos*, Greek for tyrant. Specific name from *Isan*, the local name for the northeastern part of Thailand.

Holotype. PW9-1, palaeontological collection of the Department of Mineral Resources, Bangkok, partial skeleton including the left half of the pelvis (Fig. 1), the sacrum, the 13 anteriormost caudal vertebrae (with several chevron bones), and five dorsal vertebrae.

Horizon and locality. The specimen was found in 1993 by Somchai Triamwichanon (Department of Mineral Resources, Bangkok) in red sandstones of the Sao Khua Formation at fossil site Phu Wiang 9, Phu Wiang District, Khon Kaen Province, northeastern Thailand (about 102°15' E and 16°40' N).

Age. The Sao Khu Formation, which is part of the non-marine Mesozoic Khorat Group, was long held to date from the Late Jurassic. However, palynological evidence shows that the underlying Phra Wihan Formation is Early Cretaceous in age^{8,9}. Palynomorphs and vertebrates from the Khok Kruat Formation, at the top of the Khorat Group, indicate an Aptian–Albian age^{8,10}. Consequently, the Sao Khua Formation should be considered as ante-Aptian Early Cretaceous in age. Its dinosaur assemblage¹¹ is consistent with such an age. The sauropod *Phuwiangosaurus sirindhornae*, in particular, is quite distinct from the euhelepodid sauropods from the Jurassic of China¹².

Diagnosis. A large theropod (total length ~6.5 m, by comparison with *Tyrannosaurus* which reached a length of about 12 m) with a long and relatively low ilium, the anterior blade of which forms an incipient subhorizontal medioventral shelf. Pubis with a long, straight shaft terminating in a massive distal 'boot', which is more developed anteriorly than posteriorly. No proximolateral crest on the pubis. Obturator foramen of pubis open ventrally, but largely encircled by proximal bony hook. Ischium slender, curved, with a small but well-defined scar on its proximodorsal edge for the insertion of the *Musculus flexor tibialis internus* part 3. Anterior caudal vertebrae with tall, slender neural spines. More posterior caudals with small dorsal process on the neural arch anterior to the main neural spine. Anterior chevrons long, straight and slender.

Siamotyrannus isanensis exhibits several derived features of the



FIG. 1 The left side of the pelvis of *Siamotyrannus isanensis* (holotype, Department of Mineral Resources, Bangkok, PW9-1) from the Early Cretaceous Sao Khua Formation at Phu Wiang, Thailand. Scale bar, 10 cm.

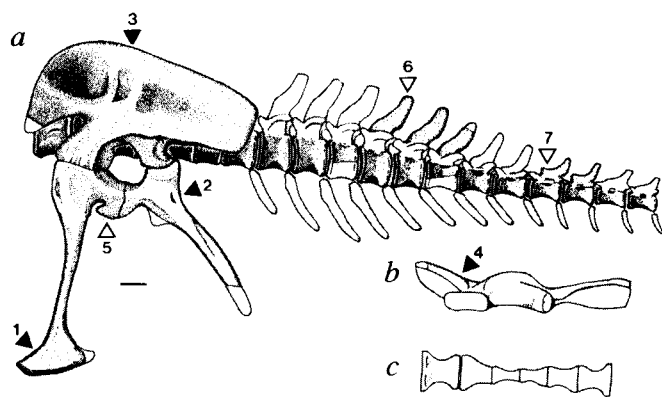


FIG. 2 a, Reconstruction of the pelvis, sacrum and anterior caudal vertebrae of *Siamotyrannus isanensis*, based on the type specimen. Unshaded parts are restored. b, Ventral view of the ilium of *Siamotyrannus isanensis*. c, Ventral view of the sacrum of *Siamotyrannus isanensis*, showing the strong lateral compression of the second and third sacral vertebrae. Black arrowheads indicate selected derived features, and white arrowheads denote primitive characters: 1, anteriorly developed distal boot of pubis; 2, muscle insertion scar on proximal part of ischium; 3, vertical ridges on iliac blade; 4, medioventral shelf of anterior iliac blade; 5, obturator notch of pubis limited by well-developed bony hook; 6, tall, slender neural spines of anterior caudal vertebrae; and 7, cranial spur on neural arch of mid-caudals. Scale bar, 10 cm. Drawings by H. Tong.

Tyrannosauridae (Fig. 2), although some of them are less fully developed than in Late Cretaceous tyrannosaurids such as *Tyrannosaurus*¹³, *Tarbosaurus*¹⁴ and *Albertosaurus*¹⁵. They include the medioventral shelf of the anterior wing of the ilium, which is absent in *Allosaurus*¹⁶, *Sinraptor*¹⁷ and other large theropods from the Jurassic but well developed in Late Cretaceous tyrannosaurids. In *Siamotyrannus*, the shelf is less horizontal and narrower than in the Late Cretaceous tyrannosaurids. The proportions of the long and relatively low ilium are like those in tyrannosaurids. In allosaurids, sinraptorids and other Jurassic theropods, the iliac blade is higher; in *Siamotyrannus*, it shows strong reliefs, including two well-marked parallel vertical ridges above the acetabulum. Similar ridges are present in Late Cretaceous tyrannosaurids, whereas Jurassic forms have smoother ilia. The medial surface of the ilium still bears parts of the broken transverse processes and neural spines of the sacral vertebrae, which were firmly attached to the ilium. The position of the insertion areas for these processes is similar to the condition in *Tyrannosaurus*¹³ and *Albertosaurus*¹⁵, with, in particular, a well-marked insertion of the transverse process of the first sacral just above the posterior end of the anterior notch, present in tyrannosaurids but not clearly visible in *Allosaurus*¹⁶ and *Sinraptor*¹⁷. The transverse process of the last dorsal fits into a depression of the medial face of the ilium, above the subhorizontal shelf, corresponding to the deep groove visible on the ilia of Late Cretaceous tyrannosaurids. Posteriorly, the medial brevis shelf is relatively narrow, as in *Allosaurus* and tyrannosaurids. The massive pubic peduncle conspicuously broadens distally, as in *Tyrannosaurus*. The prominent supra-acetabular ridge starts from the posterior part of the pubic peduncle and overhangs the ischiadic peduncle. Later tyrannosaurids have a similar ridge, often forming a strong tubercle at its posterior end that is not developed in our specimen. The pubis shows a massive distal 'boot', mainly developed anteriorly to the shaft; the posterior part may be incomplete, but was probably shorter than in Late Cretaceous tyrannosaurids. In Jurassic large theropods, the pubic boot, when present, tends to be more developed posteriorly than anteriorly (*Yuangchuanosaurus*¹⁸, *Sinraptor*¹⁷ and *Allosaurus*¹⁶). The pubic shaft of *Siamotyrannus* is straight (as in *Albertosaurus*) and more slender than in later tyrannosaurids. The pubes apparently met along part of their length distally, a condition similar to that of *Albertosaurus*. The ischium has a slender curved shaft reminiscent of Late Cretaceous tyrannosaurids, especially *Tyrannosaurus*¹³. Its anterior obturator process is broken, but a thickening of the otherwise thin anteroventral margin of the bone shows where it was located, although its exact shape cannot be determined. On the posterolateral edge of the bone, below the contact with the ilium, there is a well-marked oval scar for the insertion of the M. flexor tibialis internus part 3, which is a synapomorphy of an unnamed coelurosaurian taxon comprising the Tyrannosauridae and the Bullatosauria (ornithomimosaurs + troodontids)³. The shaft of the ischium bears a longitudinal ridge, as in *Albertosaurus*. There is no evidence of the distal thickening of the ischium seen in many Jurassic theropods (and in ornithomimosaurs) but not in tyrannosaurids, but its distal tip is missing, so there is some uncertainty in this regard. There are fewer tyrannosaurid features in the preserved vertebrae than in the pelvis. The sacrum consists of five hourglass-shaped vertebrae, with fused centra and neural spines. The second and third sacrals have very narrow centra, as in *Tyrannosaurus rex*¹³.

Siamotyrannus also exhibits features that are not found in Late Cretaceous tyrannosaurids and may be considered as primitive by comparison with other theropods (Fig. 2). One such is the presence of a well-developed obturator 'hook' on the pubis, which nearly encloses an obturator foramen. A fully enclosed foramen is present in many Jurassic theropods. In *Sinraptor*, the obturator foramen is no longer completely enclosed in bone, but there is both a proximal and a distal hook, unlike *Siamotyrannus*, which has only a proximal hook. In *Allosaurus*, the latter is much more reduced than in *Siamotyrannus*. Late Cretaceous tyranno-

saurids have practically no vestige of a pubic hook. Another apparently primitive feature for tyrannosaurids is the lack of a well-defined anterolateral ridge on the anteroproximal edge of the pubis (for the attachment of the ambiens muscle¹⁹); in *Siamotyrannus* there is merely a rugose area for muscle attachment in its place. Several probably primitive features are apparent on the vertebrae. The dorsal vertebrae (from the middle part of the dorsal series) have hourglass-shaped centra, with a dorsal depression on the lateral faces, but no pleurocoelic opening as in Late Cretaceous tyrannosaurids. The platycoelic centra of the caudals are similar to those of later tyrannosaurids, but their neural spines, when preserved, are not. Those of the fifth to seventh caudals are tall and slender, and therefore more reminiscent of *Yuangchuanosaurus*¹⁸ and *Allosaurus* than of Late Cretaceous tyrannosaurids, in which the neural spines are roughly rectangular in lateral view. More posterior caudals have lower centra with a long neural arch bearing a small median spur anterior to the main neural spine, a condition reminiscent of *Allosaurus*.

The systematic position of *Siamotyrannus* can be deduced from several of its derived features. One of these is the ischiadic scar, which characterizes a taxon within the Coelurosauria comprising the Tyrannosauridae and the Bullatosauria³. The pelvis is otherwise quite different from that of either the ornithomimosaurs²⁰, which have a different ilium with a very prominent anterior hook and a distally expanded ischium, or the troodontids²¹, which have an expanded ischium and no pubic boot. Other above-mentioned derived characters shared with Late Cretaceous tyrannosaurids also justify the inclusion of *Siamotyrannus isanensis* in the family Tyrannosauridae. Its plesiomorphic characters show, however, that it is more primitive than any of the Late Cretaceous forms. Contrary to one hypothesis¹, because of the primitive state of the obturator notch of the pubis in *Siamotyrannus*, tyrannosaurids cannot be derived from allosaurids, in which this character is more advanced. However, this would not preclude derivation from sinraptorids, in which the obturator notch is more primitive. *Siamotyrannus isanensis* is the earliest known tyrannosaurid, pushing back the record of the family by at least some 20 million years, well into the Early Cretaceous. It pre-dates all other early tyrannosaurids from Asia (*Alectrosaurus* from Inner Mongolia^{4,5}) and North America (teeth from the upper Cedar Mountain Formation of Utah⁸), which are considered as Cenomanian in age, and shows that primitive tyrannosaurids were present in Asia quite early in the Cretaceous, a finding that has biogeographical implications. All undoubted tyrannosaurids come from North America and Asia (minus India). Cretaceous theropods from India and Africa, sometimes regarded as tyrannosaurids or related to tyrannosaurids, are now placed in other families (Abelisauridae²² and Carcharodontosauridae²³, respectively). The best-known large theropod from the Lower Cretaceous of North America, *Acrocanthosaurus atokensis*²⁴, is considered as an allosaurid. This suggests that tyrannosaurids possibly evolved in Asia, at a time when the dominant large theropods of North America were allosaurids. Tyrannosaurids may thus be one of the dinosaur groups (among which ceratopsians, ornithomimosaurs and possibly hadrosaurs) that originated in Asia during a period of relative isolation in the Early Cretaceous²⁵, and later spread to North America, possibly late in the Early Cretaceous. The fauna from the Sao Khua Formation of Thailand^{11,26}, which pre-dates most Chinese and Mongolian Early Cretaceous assemblages, and includes early tyrannosaurids and ornithomimosaurs, as well as sauropods possibly related to the Late Cretaceous forms from Mongolia, is of obvious significance for our understanding of the Early Cretaceous radiation of dinosaurs in Asia. □

- Gilmore, C. W. *Bull. Am. Mus. nat. Hist.* **67**, 23–78 (1933).
- Mader, B. J. & Bradley, R. L. *J. vert. Paleont.* **9**, 41–55 (1989).
- Nessov, L. A. *Dinosaurs of Northern Eurasia: New Data about Assemblages, Ecology and Palaeobiogeography* (University of Sankt-Petersburg, Institute of Earth Crust, 1995) (in Russian).
- Kirkland, J. I. & Parrish, J. M. *J. vert. Paleont.* **15**(3) (suppl.), 39A (1995).
- Racey, A., Goodall, J. G. S., Love, M. A., Polachan, S. & Jones, P. D. in *Proc. Int. Symp. Stratigraphic Correlation of Southeast Asia* (eds Angsuwathana, P., Wongwanich, T., Tansathien, W., Wongsomsak, S. & Tulyatid, J.) 245–252 (Department of Mineral Resources, Bangkok, 1994).
- Racey, A. et al. *J. Petrol. Geol.* **19**, 5–40 (1996).
- Cappetta, H., Buffetaut, E. & Suteethorn, V. *Neues Jb. Geol. Paläont. Mh.* **11**, 659–666 (1990).
- Suteethorn, V. et al. in *Sixth Symp. Mesozoic Terrestrial Ecosystems and Biota* (eds Sun, A. & Wang, Y.) 133–136 (China Ocean Press, Beijing, 1995).
- Martin, V., Buffetaut, E. & Suteethorn, V. *C. r. Acad. Sci., Paris II* **319**, 1085–1092 (1994).
- Osborn, H. F. *Bull. Am. Mus. nat. Hist.* **35**, 733–771 (1917).
- Maleev, E. A. *Trans. Sov. Mongol. Paleont. Exp.* **1**, 132–192 (1974).
- Lambe, L. M. *Can. Geol. Surv. Mem.* **100**, 1–84 (1917).
- Madsen, J. H. *Utah Geol. Mineral. Surv. Bull.* **109**, 1–163 (1976).
- Currie, P. J. & Zhao, X. J. *Can. J. Earth Sci.* **30**, 10–11, 2037–2081 (1993).
- Dong, Z., Zhou, S. & Zhang, Y. *Palaeont. sin.* **162**, C (23), 1–145 (1983).
- Romer, A. S. *Bull. Am. Mus. nat. Hist.* **48**, 605–617 (1923).
- Barsbold, R. & Osmolska, H. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 225–248 (Univ. California Press, Berkeley, 1990).
- Russell, D. A. & Dong, Z. *Can. J. Earth Sci.* **30**, 2163–2173 (1993).
- Bonaparte, J. F. *Hist. Biol.* **5**, 1–25 (1991).
- Rauhut, O. W. M. *Berlin geowiss. Abh.* **E16**, 357–375 (1995).
- Stovall, J. W. & Langston, W. M. *Midl. Nat.* **43**, 3, 696–728 (1950).
- Russell, D. A. *Can. J. Earth Sci.* **30**, 10–11, 2002–2012 (1993).
- Buffetaut, E. et al. in *Geology, Geotechnology and Mineral Resources of Indochina* (ed. Wannakao, L.) 157–161 (Khon Kaen University, 1995).

ACKNOWLEDGEMENTS. We thank all those who took part in the excavation and preparation of the specimen, and E. Gaffney (American Museum of Natural History, New York) and R. Barsbold (Mongolian Geological Institute, Ulaan Baatar) for access to material in their care. This work was supported by the Department of Mineral Resources (Bangkok) and the French Ministry of Foreign Affairs (Paris).

CORRESPONDENCE and requests for materials should be addressed to E.B. (e-mail: buffetau@ccr.jussieu.fr).

Devastation of prey diversity by experimentally introduced predators in the field

Thomas W. Schoener & David A. Spiller

Section of Evolution and Ecology and Center for Population Biology, Storer Hall, University of California, Davis, California 95616, USA

HISTORICAL ecology contains various examples of how predators introduced onto islands by man have apparently exterminated native prey species^{1–6}. Conversely, a pioneering experiment⁷ showed an increase in number of species with predator presence. Subsequent experiments have shown both increases and decreases in prey diversity^{8–10}. Here we investigate how predator introduction affects one aspect of prey diversity (number of species or species richness), and prey abundance. We ran a seven-year experiment on an entirely natural system of small islands, using the commonest local lizard as the predator and web spiders as prey. Lizard introduction caused rapid and devastating effects on spider diversity and abundance: within two years, islands onto which lizards had been introduced became almost identical to islands with natural lizard populations. The proportion of species becoming extinct was 12.6 times higher on 'lizard-introduction' islands than on islands without lizards. Locally common and rare species were both reduced by the introduction of lizards, but nearly all of the latter became permanently extinct.

Twelve islands were selected and grouped into four similar trios according to area and vegetation characteristics. Before manipulation, each trio contained one island with lizards and two without. One of the latter two was randomly assigned to be artificially invaded with a propagule of three female and two male adult lizards, whereas the other island ('lizard-absent') was left unmanipulated.

Received 31 January; accepted 9 May 1996.

- Molnar, R. E., Kurzanov, S. M. & Dong, Z. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 169–209 (Univ. California Press, Berkeley, 1990).
- Carpenter, K. in *Aspects of Nonmarine Cretaceous Geology* (eds Mather, N. J. & Chen, P. J.) 250–268 (China Ocean Press, Beijing, 1992).
- Holtz, T. R. *J. Paleont.* **68**, 1100–1117 (1994).

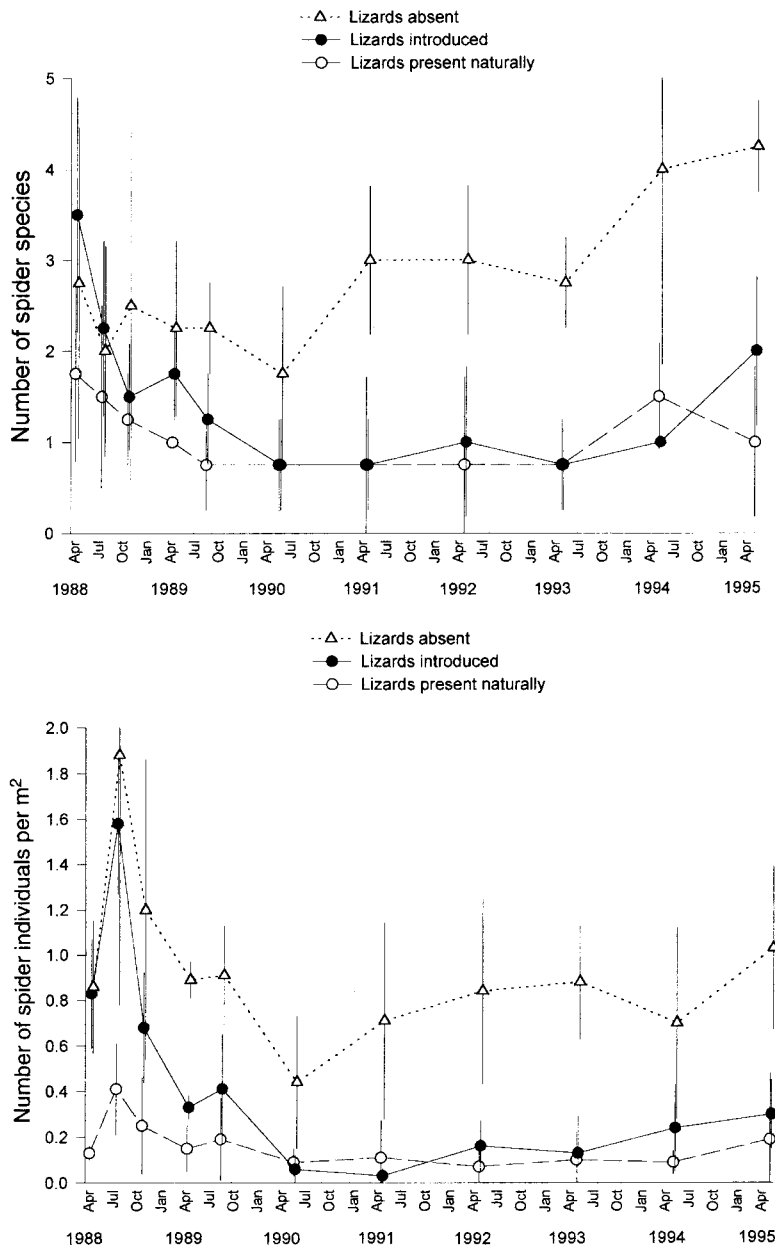
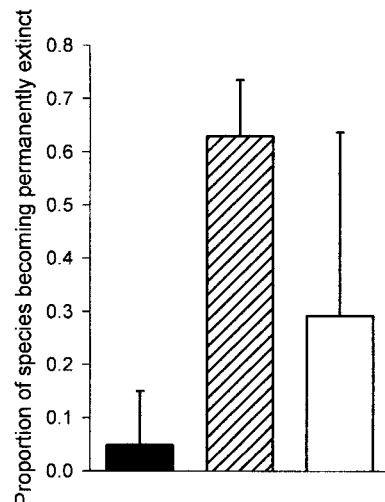


FIG. 1 Number of species (above) and densities of individuals (below) of diurnal web spiders on islands through time; leftmost values are pre-experimental. Symbols give means ($n = 4$ islands in each case); bars represent ± 1 s.d. Islands were located in the Great Abaco region of the Bahamas just south of Snake Cay; vegetated areas ranged from 40 to 179 m². All experimental islands are relatively near ($< \sim 250$ m) much larger islands and/or the mainland (these constitute the 'large-island' treatment of a more extensive experiment to be reported; see also ref. 18). They are, therefore, typical of islands with a potentially high immigration rate of species (see refs. 20, 20 for discussion of the dynamics of Bahamian metapopulation systems). Lizards used for invasion were *Anolis sagrei*, collected from the Great Abaco mainland. Mean lizard density, based on averages from 1992 and 1995 (minimum number of individuals seen) was 23% less for introduction islands than for islands with lizards present naturally, a nonsignificant difference (t -test, $P = 0.56$). Before introduction, in April 1988, we counted all individuals of diurnal web spiders on each island and classified them to species (procedures in refs 16 and 19). We did the same during post-introduction censuses performed twice at 3-month intervals during 1988, then yearly through 1995 (and also including a late-summer census in 1989). Assessing pre-introduction data, the ANOVA contrast (one-tailed) for the 2 classes of lizard-free islands compared with the lizard-present class of islands was highly significant for log density ($F = 82.52$; d.f. = 1, 9; $P = 0.0001$). The same contrast was marginally significant for number of species ($F = 2.75$; d.f. = 1, 9; $P = 0.066$). Assessing post-introduction data, the ANOVA contrast for the 2 classes of lizard islands compared with the lizard-absent class, using time averages from 1990 through 1995, was significant for both density ($F = 12.79$; d.f. = 1, 9; $P = 0.001$) and number of species ($F = 59.83$; d.f. = 1, 9; $P = 0.0001$).



FIG. 2 Mean permanent extinction (proportion of species initially present that became extinct and never recolonized) of spider species. Bars represent ± 1 s.d.. The ANOVA contrast for extinction proportion ($\sin^{-1}$ square-root transformed) for introduction compared with lizard-absent islands gives $F = 11.95$; d.f. = 1, 9; $P = 0.004$ (one-tailed test). The corresponding contrast for introduction islands compared with islands having lizards naturally gives $F = 4.33$; d.f. = 1, 9; $P = 0.0034$ (one-tailed test). The latter result is to be expected because those species most vulnerable to predation by lizards should be largely absent from islands having lizards naturally. Note that if a sequential Bonferroni test²¹ for multiple comparisons is applied to these two P values, both contrasts are still significant at $\alpha = 0.05$ for one-tailed tests or $\alpha = 0.10$ for two-tailed tests. All statistical tests were done using SAS software.



Before lizard introduction, the two classes of islands without lizards had similar spider densities of individuals and numbers of spider species (Fig. 1). These values were substantially larger than those for the islands with natural lizard populations ('lizard-present'): spider density on the former was on average 6.7 times, and the number of species was on average 1.8 times, that on lizard-present islands. After lizard introduction, both spider density and number of species dropped precipitously, and in two years the mean values of both variables nearly coincided with respective

BOX 1

Theoretical explanations for diversity-reducing or diversity-enhancing predators

See reviews in refs 5 and 10 for more extensive discussions of mechanisms.

1. Predators may decrease competition between species lower in the food web, preventing some competitor species from exterminating others, thereby enhancing diversity. The effect is strong when predators prefer dominant competitors⁷ or the most frequent prey at the time²² (for context see ref. 23). This mechanism may not occur in our system. Spider densities seem rarely high enough for severe exterminating competition to occur, perhaps because of physical factors (such as wind, driving rain, waves) affecting small organisms on small islands. Moreover, whereas habitat differences between species may reduce competition to some extent^{9,13}, habitat similarity may allow mutualistic effects, for example, the use of common webbing by heterospecifics.

2. Predators may intensify competition between species for refuges or for food resources within refuges, causing some competitor species to exterminate others, thereby reducing diversity^{5,24,25}. The mechanism generally would imply some habitat shift between the prey in the presence of predators, and it has so far been implicated in a few systems, for example freshwater fishes²⁵. No evidence exists for this mechanism in our system, although an explicit experimental test is in progress.

3. Especially abundant or otherwise effective predators may extirpate many prey individuals and perforce exterminate a substantial proportion of prey species. Indeed, even a diversity-enhancing predator can become a diversity-reducing predator if sufficiently abundant. The predator in our system, *Anolis* lizards, commonly attains densities much higher than those of other lizards²⁶. Moreover, anoles affect spiders strongly in part because both predation and competition are occurring^{14,27}. This strong negative effect reduces common species and exterminates rare ones. Furthermore, as spiders are a rather small component of the diet of Bahamian *Anolis*^{14,28}, anoles can devastate spiders and still survive on other prey such as dipterans^{5,11}. The situation is consistent with a model described in ref. 29 for 'intraguild predation' in situations of varying productivity.

4. Predators may reduce diversity when most prey species occur as small ephemeral populations renewed by immigration. A key feature of the species in our system of small islands is that they rarely occur in large populations, and natural species turnover (immigration and extinction) is high (Fig. 3 legend). Three factors may predispose such species populations to be exterminated by predators: (1) their low numbers, (2) the fact that habitat types relatively inaccessible to lizard predators are likely to be uncommon on small islands and (3) the vulnerability of their founders while searching for suitable web sites. In addition, rare species in general may be more susceptible to predation on a per capita basis (a possible relation between vulnerability and rarity in the entire system is being addressed elsewhere). The result is that species residence periods on lizard islands will be especially short; in fact, 4 species (see Fig. 3 legend) were never recorded on islands having lizards (the dramatic effect of lizard presence on spider establishment on islands has also been recently shown by field experimentation³⁰). Thus, in combination with the above three factors, a diverse species pool of immigrants can actually predispose a predator to reduce rather than enhance diversity.

means for islands having lizards naturally. The two pairs of means remained similar throughout the five remaining years of the experiment, suggesting that a new state had been reached. Moreover, our experiment showed not just that lizards reduce spider species diversity and abundance greatly (as in refs. 11–15), but that on these small islands, lizard presence and absence was a sufficient explanation for the quantitative differences between islands with and without lizards. Lizard-absent islands retained much higher values: during the last six years, spider density averaged 5.8 times, and number of species averaged 3.2 times, that on the two classes of islands with lizards.

A species is defined here as becoming permanently extinct if it disappeared from an island and did not reappear on that island for the remainder of the experiment. Most species on the lizard-introduction islands became permanently extinct: of the 14 species

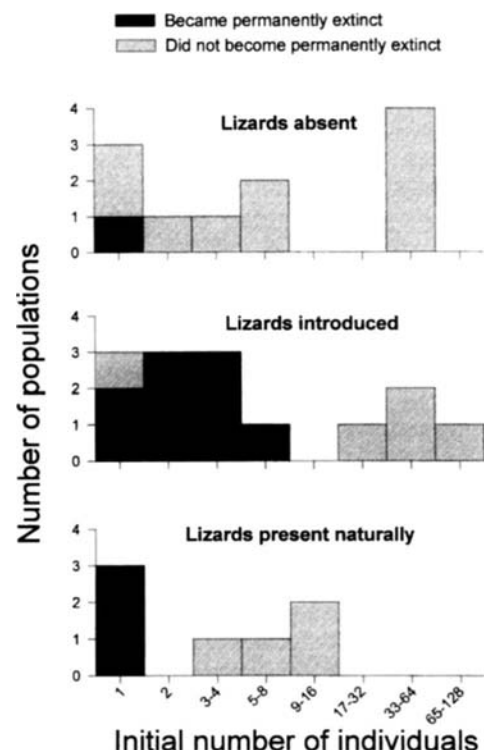


FIG. 3 Population-size distributions of species present before experimental manipulation for the three classes of islands. Each histogram combines values from all islands within the designated class. Dark-shaded portions represent populations becoming extinct and never recolonizing. A Fisher exact test for permanent extinction of rare compared with common species on introduction islands gives $P = 0.005$. The same test for the proportion of rare species that became extinct on introduction compared with lizard-absent islands gives $P = 0.004$. *Eustala cazieri* was on all 12 islands; the only common species (at least 9 individuals) was *E. cazieri*. Rare species on introduction islands becoming permanently extinct were *Argiope argentata* (3 cases), *Metepeira datona* (2 cases), *Cyclosa walckenaeri* (1 case), *Argyrodes elevatus* (1 case), *Tetragnathidae* sp. (1 case), *Uloborus trilineatus* (1 case). The last 4 species were found relatively infrequently on islands with lizards (including introduction islands after the first year); the first species, as well as *Argiope trifasciata*, *Philoponella semiplumosa* and *Metepeira* sp., never were. In addition to *E. cazieri*, *Gasteracantha cancriformis* and *Plesiometa argyra* were found relatively frequently on islands with lizards. The rare species on lizard-absent islands becoming permanently extinct was *M. datona*. Rare species on islands with natural lizard populations that became permanently extinct were *M. datona* (2 cases) and *Uloborus trilineatus* (1 case). To exemplify species turnover (see text), we calculated that on lizard-absent islands, 35 colonizations (defined as presence in year $x + 1$ but absence in year x) were recorded during the seven years of the experiment, most of which became extinct (the same figures for introduction islands and islands having lizards naturally are 13 and 8 colonizations, respectively; see also refs. 16, 19).

populations initially present on all such islands combined, 9 became permanently extinct. This compares with 1 of 11 species on the control, lizard-absent islands, and 3 of 11 on islands with natural lizard populations. Using as a measure of extinction the proportion of species initially present that became permanently extinct, the mean on lizard-introduction islands was much higher (12.6 times) than that on lizard-absent islands, and moderately higher (2.2 times) than that on islands having lizards naturally (Fig. 2).

Those species populations that became permanently extinct after lizard introduction included all but one of the ten rare populations (Fig. 3; we define, in correspondence with ref. 16, rare populations as those with fewer than nine individuals). No common (nine or more individuals) species population became permanently extinct. Only one of the seven rare species populations on lizard-absent islands became permanently extinct. Particular species could generally be characterized as either rare or common. Before introduction, all species on both classes of island without lizards except *Eustala cazieri* occurred as rare populations (however, four other species populations were sometimes common during the course of the experiment, all on lizard-absent islands after August 1988). All pre-introduction populations on islands having lizards naturally were rare except for two populations (of the four observed) of *Eustala*: populations becoming permanently extinct here were also always rare, and the proportion of species becoming permanently extinct was intermediate.

The common species *Eustala cazieri* suffered a severe reduction in numbers after lizard introduction: densities (between 1990 and 1995) on lizard-absent islands averaged 4.8 times those on the two classes of lizard islands. The same ANOVA contrasts (Fig. 1) that showed significant differences in densities of all spiders combined showed significant differences for *Eustala* alone. Nonetheless, *Eustala* remained on two lizard-introduction islands, and recolonized the other two islands after 1- and 3-year temporary extinctions. *Eustala* never became extinct on any lizard-absent island, and although it became extinct on two islands that had lizards naturally, it recolonized both. Thus *Eustala* was the only species that coexisted with lizards in the long term. In short, predator introduction usually reduced the web-spider community to a single species or eliminated it altogether, exterminating the rare species populations and greatly reducing the common ones.

A 'diversity-reducing' predator, as shown here, is perhaps intuitively expected. But because 'diversity-enhancing' predators have been found sufficiently often⁸⁻¹⁰, a full understanding of conditions favouring one or the other type is very important. Box 1 reviews theoretical explanations: in particular, the combination of small populations of prey becoming extinct and recolonizing frequently, and unusually intense predation, seems to favour diversity-reducing predators in our system.

In conclusion, we have demonstrated the devastating effect of the introduction of *Anolis* lizards on a subtropical web-spider community. Because these lizards have extraordinary proclivities for introduction¹⁷, they should be an important conservation concern for certain arthropod groups. More generally, we have used lizards and spiders to exemplify a process that, although often occurring in other groups, may be typical among terrestrial animals²: rather than preventing some species lower in the food web from competitively excluding others, thereby increasing species richness, predation exterminates rare species and reduces common ones. In such systems, predator introduction greatly threatens locally rare species, and if these are regionally rare, threatens endangered species as well. □

Received 1 November 1995; accepted 1 April 1996.

1. Diamond, J. & Case, T. J. in *Community Ecology* (eds Diamond, J. & Case, T. J.) 65–79 (Harper & Row, New York, 1986).
2. Mooney, H. A. & Drake, J. A. (eds) *Ecology of Biological Invasions* (Springer, New York, 1986).
3. Case, T. J., Bolger, D. T. & Richman, A. D. in *Conservation Biology: The Theory and Practice of Nature Conservation, Preservation and Management* (eds Fiedler, P. L. & Jain, S. K.) 91–125 (Chapman & Hall, New York, 1992).

4. Pimm, S. L. *The Balance of Nature?* (Univ. Chicago Press, Chicago, 1991).
5. Holt, R. D. & Lawton, J. H. A. *Rev. Ecol. Syst.* **25**, 495–520 (1994).
6. Simberloff, D. in *Encyclopedia of Environmental Sciences* Vol. 2 (ed. Nierenberg, W. A.) 323–336 (Academic, London, 1995).
7. Paine, R. T. *Am. Nat.* **100**, 65–75 (1966).
8. Sih, A., Crowley, P., McPeck, M., Petranka, J. & Strohmeier, J. A. *Rev. Ecol. Syst.* **16**, 269–311 (1985).
9. Fuentes, E. R. & Jaksic, F. M. *Oikos* **53**, 139–143 (1988).
10. Hixon, M. A. in *Contemporary Studies on Fish Feeding* (eds Simenstad, C. A. & Cailliet, G. M.) 235–257 (Junk Publishers, Dordrecht, 1986).
11. Spiller, D. A. & Schoener, T. W. in *Food Webs: Integration of Patterns and Dynamics* (ed. Polis, G. & Winemiller, K.) 160–169 (Chapman & Hall, London, 1996).
12. Schoener, T. W. & Spiller, D. A. *Science* **236**, 949–952 (1987).
13. Spiller, D. A. & Schoener, T. W. *Ecol. Monogr.* **58**, 51–77 (1988).
14. Spiller, D. A. & Schoener, T. W. *Oecologia* **83**, 150–161 (1990).
15. Spiller, D. A. & Schoener, T. W. *Ecology* **75**, 182–196 (1994).
16. Schoener, T. W. & Spiller, D. A. *Am. Nat.* **139**, 1176–1207 (1992).
17. Losos, J. B., Marks, J. C. & Schoener, T. W. *Oecologia* **95**, 525–532 (1993).
18. Schoener, T. W., Spiller, D. A. & Morrison, L. W. *Acta Oecologica* **16**, 103–121 (1995).
19. Schoener, T. W. & Spiller, D. A. *Nature* **330**, 474–477 (1987).
20. Schoener, T. W. *Acta Oecologica* **12**, 53–75 (1991).
21. Rice, W. R. *Evolution* **43**, 223–225 (1989).
22. Roughgarden, J. & Feldman, M. *Ecology* **56**, 489–492 (1975).
23. Schoener, T. W. *Am. Scient.* **70**, 586–595 (1982).
24. Holt, R. D. *Oikos* **50**, 276–290 (1987).
25. Mittelbach, G. G. & Chesson, P. L. in *Predation: Direct and Indirect Impacts on Aquatic Communities* (eds Kerfoot, W. C. & Sih, A.) 315–322 (Univ. Press of New England, Hanover, 1987).
26. Schoener, T. W. & Schoener, A. J. *Anim. Ecol.* **49**, 19–53 (1980).
27. Pimm, S. L. & Lawton, J. H. *Nature* **275**, 542–544 (1978).
28. Schoener, T. W. *Ecology* **49**, 704–726 (1968).
29. Polis, G. A., Myers, C. A. & Holt, R. D. A. *Rev. Ecol. Syst.* **20**, 297–330 (1989).
30. Schoener, T. W. & Spiller, D. A. *Science* **267**, 1811–1813 (1995).

ACKNOWLEDGEMENTS. We thank H. Levi and B. Opell for species identification, and J. Diamond, M. Hixon, R. Holt and G. Mittelbach for comments on a previous draft. Supported by grants from the NSF.

CORRESPONDENCE and requests for materials should be addressed to T.W.S. (e-mail: twschoener@ucdavis.edu).

Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*

Travis T. Kibota* & Michael Lynch†

* Biology Department, Clark College, 1800 E. McLoughlin Blvd, Vancouver, Washington 98663, USA

† Department of Biology, University of Oregon, Eugene, Oregon 97403, USA

Mutations are a double-edged sword: they are the ultimate source of genetic variation upon which evolution depends, yet most mutations affecting fitness (viability and reproductive success) appear to be harmful¹. Deleterious mutations of small effect can escape natural selection, and should accumulate in small populations^{2–4}. Reduced fitness from deleterious-mutation accumulation may be important in the evolution of sex^{5–7}, mate choice^{8,9}, and diploid life-cycles¹⁰, and in the extinction of small populations^{11,12}. Few empirical data exist, however. Minimum estimates of the genomic deleterious-mutation rate for viability in *Drosophila melanogaster* are surprisingly high^{1,13,14}, leading to the conjecture that the rate for total fitness could exceed 1.0 mutation per individual per generation^{5,6}. Here we use *Escherichia coli* to provide an estimate of the genomic deleterious-mutation rate for total fitness in a microbe. We estimate that the per-microbe rate of deleterious mutations is in excess of 0.0002.

Fifty genetically identical *E. coli* lines were maintained for 300 days by picking and streaking of single colonies per line (Fig. 1). Lines were maintained in an environmental chamber under fluorescent lighting at a constant 37 °C. Selection against deleterious mutations was reduced because of repeated population bottlenecks (see later). This allowed an accumulation of deleterious mutations of small effect (those deleterious mutations that are evolutionarily relevant), with independently maintained lines accumulating different numbers and types of mutations. Two results were expected: (1) mean fitness (exponential growth

rate; Fig. 2) should have decreased over time; (2) divergence should have occurred, leading to increased among-line variance for fitness. The expected patterns were obtained (Fig. 3).

Mean fitness decreased linearly over time, and the among-line variance of fitness increased linearly. Estimation of the rates of these changes leads to a lower-bound estimate of the genomic deleterious mutation rate ($U_{\min}$) and an upper-bound estimate of

the average deleterious mutational effect ($\bar{s}_{\max}$; Fig. 3). Our data indicate $\bar{s}_{\max} = 0.012$ as the decrease in fitness per mutation. This is similar to the $\bar{s}_{\max}$ estimate of *Drosophila* (0.027) (refs 1,13). The bias in these estimates arises because of a lack of information on θ_s , the squared coefficient of variation of mutational effects (Fig. 3). Direct estimates of θ_s are not available for any organism, but mutational effects in *Drosophila* appear to be highly dispersed.

FIG. 1 Representation of the mutation accumulation phase of the experiment.

METHODS. Fifty lines of an *E. coli* B strain (see ref. 17 for strain description and procedures for contamination checks) were isolated by picking a single colony (the founder colony) from an agar plate, streaking onto a new plate, incubating overnight at 37 °C, then repeating the pick-streak-incubate procedure (called a growth cycle, GC) for 50 of the resultant colonies. These 50 lines were grown under identical conditions for 300 growth cycles. All cultures were maintained on Davis minimal (DM) agar plates²² (supplemented with 0.2 g glucose and 2×10^{-6} g thiamine hydrochloride per litre). Preliminary investigation using mixed cultures of the founder strain and a derived colour variant¹⁷ showed that colonies result from single cells under these conditions. Thus, all fifty lines were derived from a single cell (the founding cell of the founder colony), and were genetically identical except for the mutations that might have occurred during colony expansion (growth cycle 0). Throughout the mutation accumulation phase, there were ~25 rounds of cell division per growth cycle. Thus each line experienced 300 single cell bottlenecks and ~7,500 rounds of cell division. At growth cycles 0, 100, 120, 200, 250 and 300, sub-samples of each line were frozen in a glycerol-based suspension at -80 °C. Under these conditions, cells can be maintained indefinitely without evolution¹⁷. These frozen samples were stored for use in fitness assays.

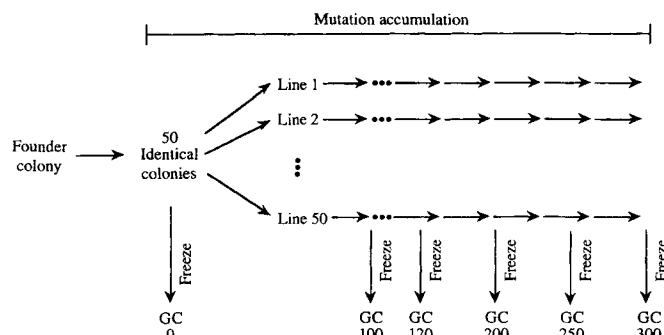


FIG. 2 Time series of $\log_e$ of optical density for a typical assay tube. Lines were assayed in liquid cultures (4 ml DM with 0.15 g glucose and 2×10^{-6} g thiamine hydrochloride per litre in 13-mm test tubes) at 37 °C and shaken at 200 r.p.m. Thawed sub-samples of frozen material were incubated in fresh assay medium on two consecutive nights before being assayed. These two preliminary incubations minimized freeze-thaw effects. Cells (40 μ l) were transferred to fresh tubes and incubated at 37 °C and 200 r.p.m. These tubes were repeatedly measured for optical density (at 400 nm; Milton-Roy Spectronic 21-D) to obtain growth curves. Optical densities were log-transformed to linearize the exponential growth phase. The slope of this linear region yields an estimate of exponential growth, r , which incorporates survival and reproductive rate and is therefore a measure of total fitness. The slope [$f'(t)$] was evaluated at the inflection point (t^* , y^*) of the fitted cubic regression, $f(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3$. The inflection point (where $f''(t) = 0$) was chosen because it is in the linear (middle) region of these curves. In some cases (619 out of 2,350), too few points were obtained in stationary phase to produce a significant α_3 coefficient. These cases were analysed using quadratic regressions, $f(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$. No inflection point exists in these curves. The time point at which the slope was evaluated was therefore chosen by substituting the mean y^* (from the controls) for $f(t)$ and solving for t^* . Because the middle regions of the curves are approximately linear, estimations of the slopes are not very sensitive to t^* ; mean fitness from cubic regressions (0.982) did not differ significantly from the mean of quadratic regressions (0.987; $P = 0.24$ by a t -test).

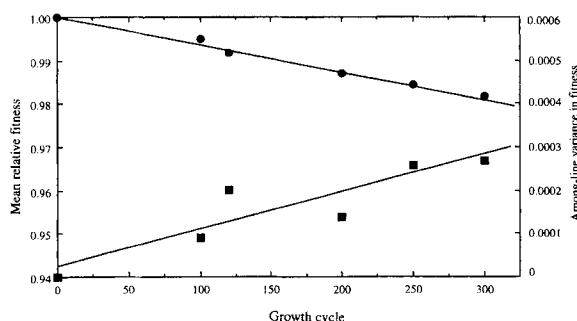
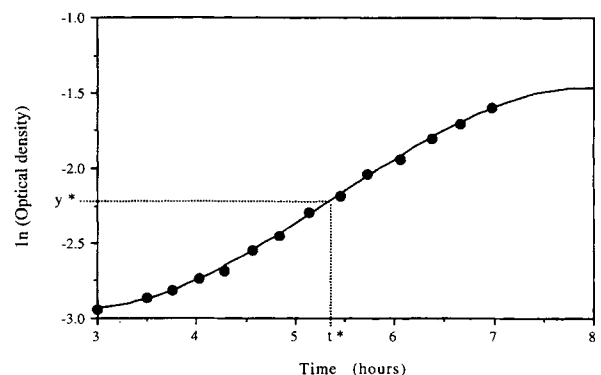


FIG. 3 Mean fitness (circles) and the among-line variance for fitness of the 50 lines (squares) over time. Lines were assayed in randomized blocks containing one replicate from each line for a single growth cycle (GC) cohort, and either 25 random or all 50 lines form the control (GC 0) cohort. Fitness was estimated as the exponential of growth rate (e^r). Analysis of variance revealed significant among-block effects, but no significant (block $\times$ GC) or (block $\times$ line) interactions. Block effects were removed by standardizing with the mean control exponential of growth rate (within blocks). Assuming a multiplicative fitness model (no epistasis), the expectation of log relative fitness after t generations is $\ln(W_t) = U t E[\ln(1 - s)] \approx -U s t$, where U is the mutation rate per haploid genome per generation, s is the fractional reduction of fitness due to a mutation, and E denotes the expected value of the quantity in brackets. After t generations, the expected variance for log relative fitness among lines is approximately $\sigma^2[\ln(W_t)] = U E(s^2) t$. Setting $R = U s$ and $V = U E(s^2)$, and rearranging, yields $\frac{R^2}{V} = \frac{U}{1 + \theta_s}$ and $\frac{V}{R} = \frac{s}{1 + \theta_s}$, where $\theta_s = \frac{s^2}{s}$ is the squared coefficient of variation of mutational effects. As $\theta_s \geq 0$, $\frac{R^2}{V}$ provides a lower bound for the genomic mutation rate ($U_{\min}$) and $\frac{V}{R}$ provides an upper bound for the average mutational effect $\bar{s}_{\max}$ (refs 7,23). Estimators for these boundary values are $U_{\min} = \frac{R^2(1 - C_R)}{V(1 + C_V/2)}$ and $\bar{s}_{\max} = \frac{V}{R(1 + C_R)}$, where C_R and C_V are the squared coefficients of sampling variation of the estimates R and V (ref. 24).

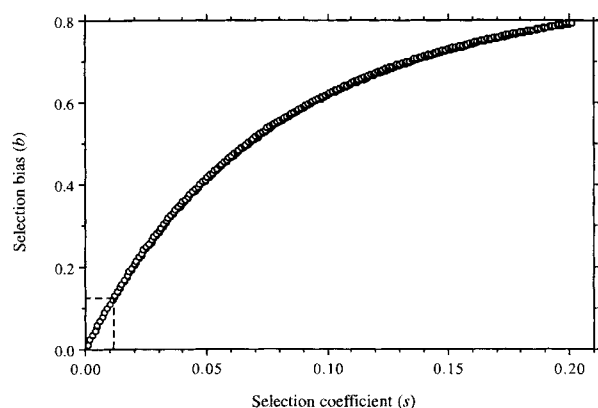


FIG. 4 The probability that a newly arisen deleterious mutation is lost due to selection during colony growth. Selection bias values (b) were obtained from a deterministic model of colony growth that assumes that a single cell is transferred after 24 h of colony growth and that newly mutated cells grow exponentially at the rate r_M whereas their immediate ancestors (colony founders) grow at the rate r_0 . Letting N_0 and N_M be the numbers of cells with founder genotypes and with new mutations during any growth cycle, the initial colony state is $N_0(0) = 1$ and $N_M(0) = 0$, and the number of non-mutant and mutant cells at time t are $N_0(t) = e^{(r_0 - U)t}$ and $N_M(t) = [U(e^{(r_0 - U)t} - e^{(r_M - U)t})]/(r_0 - r_M - U)$. Letting $N(t)$ be the total number of cells in a colony, the fraction of cells that are new mutants at the time of cell passage is simply $N_M(t)/N(t)$ for $s = (r_0 - r_M)$, which is approximately equal to $1 - e^{-Ut} \approx Ut$ when mutations are neutral. Plotted values give $b = [1 - N_M(t)/N(t)]$ for $s = (r_0 - r_M)$ relative to the neutral case. Results are given for genomic deleterious-mutation rates of $U = 0.001$ (solid points), 0.0001 and 0.00001 per hour (open points). Dashed lines denote the selection bias, given the average deleterious effect estimated in this study (~ 0.012). This is a deterministic model—stochasticity (random genetic drift) should cause a net decrease in the efficiency of selection. Thus, this model produces an upper bound on the selection bias. U_s , the lower-bound mutation rate estimate adjusted for selection bias, is obtained as $U_{\min}(1 + b)$. We estimate $b \approx 0.13$, and $U_s \approx 1.9 \times 10^{-4}$ mutations per genome per generation.

If the distribution of mutational effects of *E. coli* is similar in form to that of *Drosophila*, then $\bar{s}_{\max}$ could overestimate the true value by as much as an order of magnitude. Hartl *et al.*¹⁵ have derived indirect estimates of $\bar{s}$ for natural populations of *E. coli* and *Salmonella typhimurium*. They found $\bar{s}_{\max} = 2 \times 10^{-8}$ for non-synonymous base substitutions (those causing amino-acid changes) and 7×10^{-9} for synonymous base substitutions (those not causing amino-acid changes). Thus, deleterious mutations observed in natural populations appear to have selection coefficients several orders of magnitude lower than the mutations in the present experiment. This is expected because mutations of large effect should be quickly eliminated by selection in typically large natural populations (those not experiencing repeated bottlenecks).

The mutation rate estimate from the regressions is $U_{\min} = 1.7 \times 10^{-4}$ mutations per genome per generation. This estimate may be downwardly biased by selection against deleterious mutations during colony growth. Figure 4 shows the results of a deterministic model of selection that puts an upper limit on this second source of bias. U_s , the lower-bound estimate of the mutation rate, adjusted for selection bias, is estimated to be 1.9×10^{-4} mutations per genome per generation (Fig. 4). The true deleterious-mutation rate may be somewhat higher for several reasons: (1) θ_s is greater than zero (see above); (2) only a fraction of the mutations that accumulated was expressed (for example, mutations at catabolic operons such as *lac* and *mal* could not be detected); and (3) an unknown fraction of mutations may have been selected out during stationary phase of the growth cycles. This estimate is an order of magnitude lower than the total genomic mutation rate for *E. coli* ($U_{\text{total}} = 3 \times 10^{-3}$) estimated by Drake¹⁶.

An assumption in our experiment is that all mutations affecting fitness were deleterious. The true deleterious-mutation rate would be underestimated if beneficial mutations occurred at appreciable rates. Lenski *et al.*^{17,18} used the parental strains of our lines for experiments similar to ours in their growth conditions. They studied adaptive mutations by maintaining populations at a large size (between 10^6 and 10^8 cells). During the first year ($\sim 2,000$ generations), they observed the fixation of very few adaptive mutants ($U = 10^{-9}$ to 10^{-10} mutations per cell per generation), and during the subsequent four years observed an even lower rate of adaptation. In our experiment, population sizes were considerably smaller in all but the last few cell divisions of each growth cycle. This combination of low rate of adaptive mutation (even under intense selection)^{17,18} and the population bottlenecks that we maintained should have led to a negligible rate of fixation of beneficial mutations. The downward bias introduced by the assumption of only deleterious mutations is therefore not likely to be great. This low rate of beneficial mutation also makes unlikely the fixation of mutations that were beneficial on the agar medium used in mutation accumulation but pleiotropically deleterious in the liquid assay medium.

The genomic deleterious-mutation rate of *E. coli* is several-hundred-fold lower than that of *Drosophila* ($U_{\min} \approx 0.3$ deleterious mutations per haploid genome per generation)^{1,13}. Possible reasons for this difference include a 50-fold difference in haploid genome size (4×10^6 base pairs in *E. coli*¹⁹ versus 200×10^6 base pairs in *Drosophila*²⁰) and a 25-fold difference in the number of cell divisions per generation (one for *E. coli* versus about 25 germ-line divisions for *Drosophila*²¹). Assuming that the genomic mutation rate is proportional to genome size and the number of cell divisions per generation, extrapolation of the deleterious genomic mutation rate for *E. coli* to *D. melanogaster* becomes $(1.7 \times 10^{-4}) \times 50 \times 25 = 0.21$ mutations per haploid genome per generation. The general agreement (to an order of magnitude) between the estimates for these organisms is compelling. □

Received 22 February; accepted 24 April 1996.

1. Crow, J. F. & Simmons, M. J. in *The Genetics and Biology of Drosophila* Vol. 3C (eds Ashburner, M., Carson, H. L. & Thompson, J. N.) 1–35 (Academic, London, 1983).
2. Charlesworth, D., Morgan, M. T. & Charlesworth, B. *Genet. Res. Camb.* **61**, 39–56 (1993).
3. Haigh, J. *Theor. Pop. Biol.* **14**, 251–267 (1978).
4. Lynch, M. & Gabriel, W. *Evolution* **44**, 1725–1737 (1990).
5. Muller, H. J. *Mut. Res.* **1**, 1–9 (1964).
6. Kondrashov, A. S. *Nature* **336**, 435–440 (1988).
7. Charlesworth, B. *Genet. Res. Camb.* **55**, 199–221 (1990).
8. Charlesworth, B. in *Sexual Selection: Testing the Alternatives* (eds Bradbury, J. W. & Andersson, M. B.) 21–40 (Wiley, Chichester, 1987).
9. Kirkpatrick, M. & Ryan, M. J. *Nature* **350**, 33–38 (1991).
10. Jenkins, C. D. & Kirkpatrick, M. *Evolution* **49**, 512–520 (1995).
11. Lynch, M., Burger, R., Butcher, D. & Gabriel, W. *J. Heredity* **84**, 339–344 (1993).
12. Lande, R. *Evolution* **48**, 1460–1469 (1994).
13. Mukai, T. in *Quantitative Genetic Variation* (eds Thompson, J. N. & Thoday, J. M.) 177–196 (Academic, New York, 1979).
14. Keightley, P. D. *Genetics* **138**, 1315–1322 (1994).
15. Hartl, D. L., Moriyama, E. N. & Sawyer, S. A. *Genetics* **138**, 227–234 (1994).
16. Drake, J. W. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7160–7164 (1991).

17. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. *Am. Nat.* **138**, 1315–1341 (1991).
18. Lenski, R. E. & Travisano, M. *Proc. natn. Acad. Sci. U.S.A.* **91**, 6808–6814 (1994).
19. Herdman, M. in *The Evolution of Genome Size* (ed. Cavalier-Smith, T.) 37–68 (Wiley, London, 1985).
20. Cavalier-Smith, T. in *The Evolution of Genome Size* (ed. Cavalier-Smith, T.) 69–104 (Wiley, London, 1985).
21. Lindsley, D. L. & Tokuyasu, K. T. in *The Genetics and Biology of Drosophila* (eds Ashburner, M. & Wright, T.) 226–294 (Academic, New York, 1980).
22. Carlton, B. C. & Brown, B. J. in *Manual of Methods for General Bacteriology* (ed. Gerhardt, P.) 222–242 (Am. Soc. Microbiol., Washington DC, 1981).
23. Bateman, A. J. *Int. J. Rad. Biol.* **170**–180 (1959).
24. Lynch, M. in *Ecological Genetics* (ed. Real, L.) 86–108 (Princeton Univ. Press, Princeton, New Jersey, 1994).

ACKNOWLEDGEMENTS. We thank R. Lenski for bacterial strains, advice and encouragement; M. Teiser for laboratory assistance and for microbiological consultation; H.-W. Deng for statistical advice; R. Lande, F. Stahl, L. Chao and T. Leatham for discussion; and N. Lehman, D. Price, D. Butcher, M. Pfender, M. Jordan, R. Myers, R. Lenski and J. Peck for comments on earlier versions of this manuscript. This work was supported by the NIH and the T&EK Foundation.

CORRESPONDENCE and requests for materials should be addressed to T.T.K. (e-mail: tkibota@clark.edu).

Visual feature selectivity in frontal eye fields induced by experience in mature macaques

Narcisse P. Bichot, Jeffrey D. Schall
& Kirk G. Thompson

Vanderbilt Vision Research Center, Department of Psychology,
Vanderbilt University, Nashville, Tennessee 37240, USA

WHEN examining a complex image, the eye movements of expert observers differ from those of novices; experts have learned to ignore features that are visually salient but are not relevant to the interpretation of the image¹⁻³. We have studied the neural basis of this form of perceptual-motor learning using monkeys that have learned to search for a visual target among distractors. Monkeys trained to search only for, say, a red stimulus among green distractors will ignore green stimuli even if they subsequently appear as targets in a complementary search array, that is, among red distractors. We recorded from neurons in the frontal eye field (FEF), a cortical area that responds to visual stimuli and controls purposive eye movements⁴⁻⁶. Normally, FEF neurons do not exhibit feature selectivity, but their activity evolves to signal the target for an incipient eye movement⁷. In monkeys trained exclusively on targets of one colour, however, FEF neurons show selectivity for stimuli of that colour. Because this selective response occurs so soon after presentation of the stimulus array, and is independent of location within the visual field, we propose that it reflects a form of experience-dependent plasticity that mediates the learning of arbitrary stimulus-response associations.

To investigate the effects of training experience on gaze behaviour and the underlying neural activity, a between-subjects experimental design was used. Control monkeys were trained to make saccades to the single item of one colour among an array of elements of a different colour, in two complementary search arrays (red target among green distractors, and green target among red distractors). Experimental monkeys were given exclusive experience with only one of the two complementary search arrays. Control monkeys shift their gaze according to visual salience (Fig. 1, left column), but experimental monkeys persistently direct gaze to stimuli possessing a specific colour (right column). Despite this difference, the saccade latencies of experimental monkeys were not different from those of control monkeys (experimental, mean 199 ms, s.e.m. 0.5 ms; control, mean 198 ms, s.e.m. 0.8 ms; $t_{9386} = 1.40$). Similar experiments using humans have demonstrated a significant increase in response times when the colours of the target and distractors reverse unpredictably across trials compared to when they remain constant⁸. No such difference was observed in our data, probably because we changed the target and distractor colours across, but not within, blocks of trials with the control monkeys.

The FEF, located in the rostral bank of the arcuate sulcus in the frontal cortex, receives topographic, convergent input from visual areas responsible for form and object processing, as well as from areas involved in motion and spatial vision⁹. The FEF also projects strongly onto the subcortical structures that control the generation of eye movements, and stimulation of FEF with low currents evokes saccades⁶. Therefore, FEF is uniquely positioned to play a central role in the generation of a command to shift gaze, based on the product of visual processing⁴⁻⁶. Different classes of neurons in the FEF responded to visual stimuli or discharge in association with eye movements^{10,11}. We recorded from 80 neurons that responded to visual stimuli in 90 penetrations from the FEF in

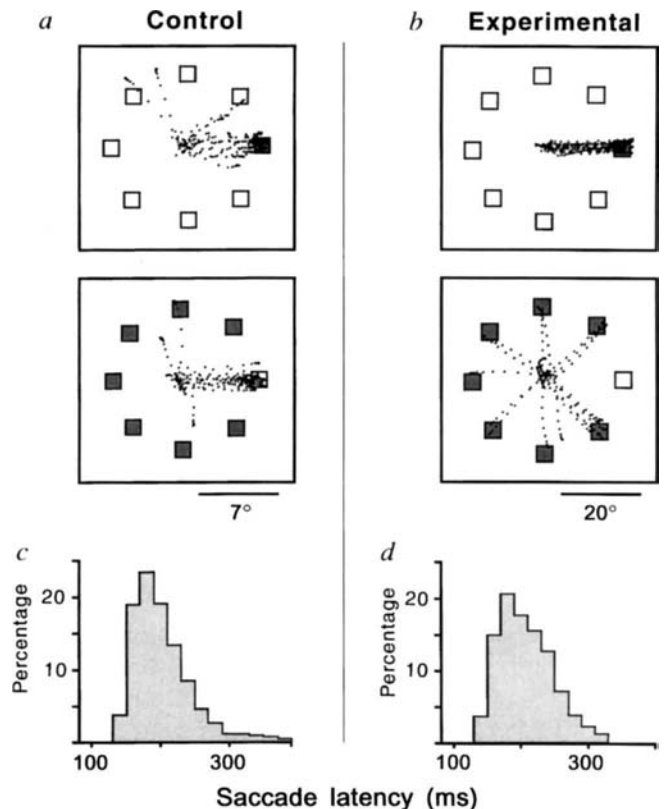


FIG. 1 Effects of experience on eye-movement responses to visual search arrays. Data were collected from 4 macaques, *Macaca mulatta*, of mass 4–10 kg. The behavioural and physiological methods have been described⁷. Each trial began when the monkey fixated a central white spot on a video monitor (60 Hz refresh rate). After an interval of fixation (400–500 ms) the target was presented either alone or with distractors. Within each block the target was presented at one of eight positions varying randomly in direction at each cell's optimal eccentricity. A trigger signal was given by a colour change of the fixation spot (simultaneous with stimulus presentation) to make a saccade to the target. If the monkey failed to direct its gaze properly at any time, the trial was aborted and no reinforcement was given. Targets and distractors were distinguished by colour (red versus green). Stimuli were presented on a grey background and were adjusted to be isoluminant. Saccades were detected using a previously described algorithm⁷. The key manipulation was that experimental monkeys were trained exclusively with one visual search array. One experimental monkey only experienced a green target among red distractors, and the other experimental monkey only experienced a red target among green distractors. The two monkeys were trained on each of the complementary search arrays to rule out possible stimulus confounds. *a*, Gaze behaviour of a control monkey trained with both complements of a search stimulus array. The target is shown at the right horizontal position. The first saccade was made to fixate the red target among green distractors (top) or the green target among red distractors (bottom) except for infrequent errors. *b*, Gaze behaviour of an experimental monkey. In response to the over-trained search array the monkey made saccades directly to the salient red target (top). When presented with the complementary search array of a green target among red distractors (bottom), the monkey shifted gaze to the distractors and not to the target. These trials were errors, and no reward was given. *c*, *d*, Distribution of visual search saccade latencies for all trials collected during recordings from all cells. The gaze behaviour of experimental monkeys in response to the unlearned stimulus array was probed with a few trials every week to insure the stability of the behavioural response bias. We did not expose the monkeys to the unlearned array for too many trials to prevent them from learning to shift gaze to the salient stimulus, thereby losing the behavioural response bias. Future studies will investigate how visual responsiveness in FEF changes as monkeys learn to generalize across search arrays.

the two experimental monkeys; 47 neurons had visually evoked activity and provided sufficient data for this report. These data were compared to corresponding unit activity in the control monkeys⁷. The activity of representative cells from each of the experimental monkeys is shown in Fig. 2. In striking contrast to our findings in the control monkeys, the initial visual response to the distractors of the search array was significantly less than the initial visual response to the target.

The ratios of 'initial visual response to target' to 'initial visual

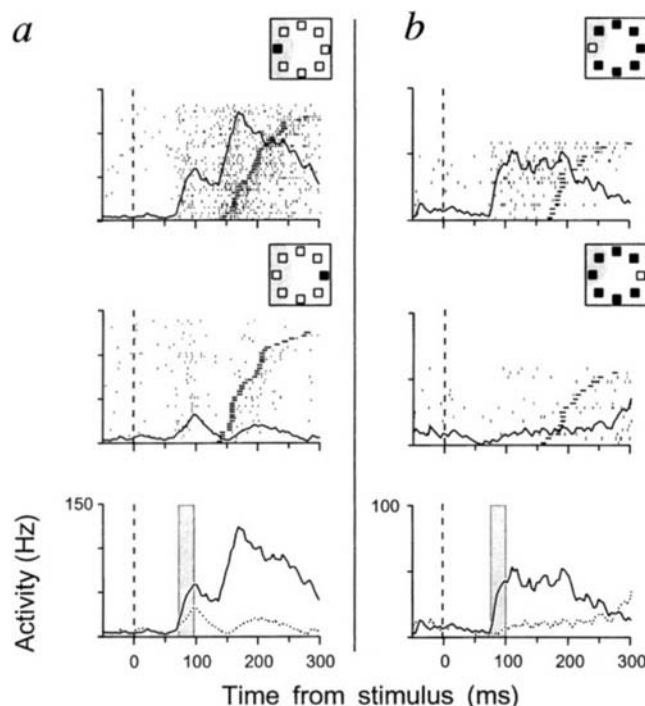


FIG. 2 Apparent colour-selective visual response of FEF neurons in experimental monkeys. Neural activity is represented by a raster plot; each vertical tick mark indicates the time of an action potential, the horizontal line in each raster indicates the time of saccade initiation, and rasters are ordered by saccade latency. Superimposed on the raster is the average spike density function; the ordinate scale represents the discharge rate. The rasters and spike density function are aligned on the time of stimulus presentation at time zero (vertical broken lines). Visual responses were identified by their consistent latency relative to the time of stimulus appearance. Stimulus configuration in relation to the cell's response field (shaded) is indicated. Responses of cells when the target was in their receptive field (top) were compared with responses to only distractors in their receptive field (middle). The average spike density function from these two conditions are superimposed (bottom; solid line, target in receptive field; dotted line, only distractors in receptive field). We analysed all neurons that responded to the visual search stimulus array at eccentricities ranging from 4° to 20° . To determine whether the initial response of cells discriminated whether the target or a distractor was in the response field, we counted the spikes in the first 25 ms of visually evoked activation (bottom, shaded area). The resulting spike count values for all trials were submitted to an analysis of the relative variances using an *F*-test. Then, depending on the outcome, we applied a two-tailed *t*-test for equal or unequal variances to test for differences in the response to the target of the search array versus the distractors in the receptive field. This analysis depended on accurately determining the visual response latency of the cell. Application of a spike-train analysis that detects periods of significantly elevated activity in single trials to the visual responses of cells allowed an accurate estimation of the visual response latency²⁴. Visual latency estimates were corrected for the raster scan time based on the location of the receptive field on the video monitor; average correction, 7.4 ms. For both of these neurons, the initial visual response was significantly greater when the target was in the response field than when only distractors were in the response field in the 25 ms after the beginning of the visual response (a, $t_{80} = 3.38$, $P < 0.01$; b, $t_{23} = 4.83$, $P < 0.01$).

response to distractors' for all cells from the experimental monkeys were significantly greater than the comparable ratios for control monkeys (experimental response ratio, mean 1.70, s.e.m. 0.17, range 0.61–6.50; control response ratio, mean 1.17, s.e.m. 0.04, range 0.62–1.92; Mann–Whitney *U* test, $U = 631.0$, $P < 0.01$). Figure 3 plots the probability that the initial visual responses to the target were the same as the responses to distractors, as a function of visual response latency for the cells from control (open circles) and experimental (filled circles) monkeys. In control monkeys there was no systematic tendency for cells to respond preferentially to the target or the distractor. However, in each case of initial visual discrimination in experimental monkeys (in 21 of 47 neurons), the initial visual response to the target was greater than the response to distractors. The magnitude of response of the colour-selective cells evoked by the target was no different from the magnitude of response of the non-selective cells. Instead, the visual selectivity occurred because of an attenuated and/or delayed response to the distractor. The sample of colour-selective cells in the FEF had receptive fields representing all parts of the visual field from 4° to 20° eccentricity.

The visual latencies of the subset of cells from the experimental monkeys that discriminated the target from distractors in their initial response ranged from 50 to 129 ms, with a mean of 79 ms. The incidence of cells exhibiting an initial target-selective visual response was significantly higher in experimental monkeys in both early (<70 ms; Fisher exact test, $P = 0.01$) and late (>70 ms; Fisher exact test, $P = 0.02$) ranges of visual response latencies. The distribution of visual response latencies for all cells recorded in control monkeys was significantly earlier than the distribution for all cells recorded in experimental monkeys (control, mean 60 ms, s.e.m. 3.5; experimental, mean 71 ms, s.e.m. 3.0; Mann–Whitney *U* test, $U = 675.0$, $P < 0.01$).

The characteristic differences in gaze behaviour observed in the experimental and control monkeys indicate that different behavioural search strategies arose through experience with either constant or changing visual search arrays. By overtraining mon-

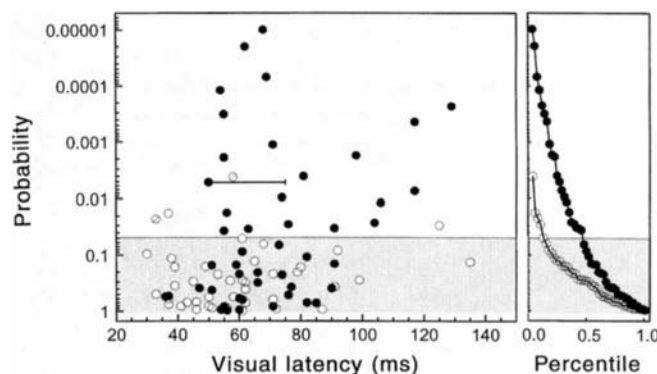


FIG. 3 Statistical analysis of the initial visual activation of neurons recorded in control (open circles) and experimental (filled circles) monkeys. The probability of the initial response to the target in the response field being the same as the initial response to the distractors in the response field is plotted as a function of the visual response latency (left). The shaded region indicates nonsignificant probability values greater than 0.05. Of the 43 neurons from control monkeys, 39 fell in the nonsignificant area, two responded preferentially to the target, and two responded preferentially to the distractors of the search array field (marked by diagonal lines). In contrast, 21 of 47 neurons recorded from the experimental monkeys exhibited significantly greater initial responses when the search array target fell in the response field, and none showed the opposite effect. The greater magnitude and incidence of the early visual discrimination observed in experimental as compared with control monkeys is highlighted by the plot of the cumulative distributions of the probabilities (right). The earliest response latency at which a cell from an experimental monkey exhibited a significant difference was 50 ms, representing an interval of analysis extending to 75 ms (indicated by the horizontal bar to the right of the point).

keys on one search array, we created a condition in which FEF cells express a visual stimulus selectivity they would not otherwise have^{7,12,13}. Overtraining also led to an apparent increase in the distribution of visual response latencies of FEF neurons. Further work is needed to establish whether the latencies of individual neurons change, or whether different populations of neurons have been recorded in the control and experimental monkeys.

Our finding is probably related to the enhancement of the visual responses of neurons in visuomotor structures including FEF observed specifically when the stimulus in a neuron's receptive field is used as the target for a gaze shift^{13,14}. However, unlike earlier studies that presented one or two stimuli in blocks of trials, in our experiment distractors were always present, and target location was much less predictable. Therefore, the early visual selectivity we observed in the experimental monkeys is unlikely to be due only to directed attention or motor planning specific for a particular visual field location.

Discrimination of a target from distractors based on visual salience or the subject's instructed preference has been observed in visual cortex¹⁵⁻¹⁹. This discrimination occurs typically 140–150 ms after stimulus presentation, a time which coincides with the time of visual target discrimination by FEF neurons measured in control monkeys²⁰. The latency of the colour based discrimination observed in FEF of the experimental monkeys was markedly shorter than the attention-related modulations observed in extrastriate visual cortex. Indeed, the timecourse of the early visual discrimination in FEF coincides with or shortly follows the latency of activation of colour-selective cells in macaque primary visual cortex, having estimated mean values ranging from 45 ms (ref. 21) to 80 ms (ref. 22). If there is insufficient time for attentional modulation based on stimulus evaluation, it is possible that the attenuated response of FEF neurons to the distractors is mediated by a reduction in the synaptic efficacy of neurons representing the constant distractor feature. The present data do not show whether the hypothetical synaptic plasticity occurs in FEF or in visual areas that register the colour of the search stimuli.

If plasticity underlies the FEF visual selectivity observed here, then this finding contrasts with previous work, because it demonstrates a change in neural selectivity due to selective experience that was not localized in a topographic map. Previous reports of neural plasticity based on experience in adult primates have described expansions of representations within topographic maps that are associated with perceptual or motor skill acquisition²³. The experience-dependent, early visual selectivity we observed in FEF was not due to a topographically limited pattern of sensory stimulation or motor output, but rather was contingent on a particular stimulus–response mapping. This finding may indicate another form of adaptation of the mature brain that is associated with the establishment of a habit, if not a skill. □

Received 30 January; accepted 8 May 1996.

1. Fisher, D. F., Monty, R. A. & Senders, J. W. *Eye Movements: Cognition and Visual Perception* (Erlbaum, Hillsdale, NJ, 1981).
2. Groner, R., McConkie, G. W. & Menz, C. *Eye Movements and Human Information Processing* (Elsevier, New York, 1985).
3. O'Regan, J. K. & Levy-Schoen, A. *Eye Movements: From Physiology to Cognition* (Elsevier, New York, 1987).
4. Goldberg, M. E. & Segraves, M. A. in *The Neurobiology of Saccadic Eye Movements* (eds Wurtz, R. H. & Goldberg, M. E.) 283–313 (Elsevier, New York, 1989).
5. Bruce, C. J. in *Signals and Sense, Local and Global Order in Perceptual Maps* (eds Edelman, G. M., Gail, W. E. & Cowan, W. M.) 261–314 (Wiley, New York, 1990).
6. Schall, J. D. in *The Neural Basis of Visual Function* (ed. Leventhal, A. G.) 388–442 (Macmillan, London, 1991).
7. Schall, J. D., Hanes, D. P., Thompson, K. G. & King, D. J. *J. Neurosci.* **15**, 6905–6918 (1995).
8. Bravo, M. J. & Nakayama, K. *Percept. Psychophys.* **51**, 465–472 (1992).
9. Schall, J. D., Morel, A., King, D. J. & Bullier, J. *J. Neurosci.* **15**, 4464–4487 (1995).
10. Bruce, C. J. & Goldberg, M. E. *J. Neurophysiol.* **53**, 603–635 (1985).
11. Schall, J. D. *J. Neurophysiol.* **66**, 559–579 (1991).
12. Mohler, C. W., Goldberg, M. E. & Wurtz, R. H. *Brain Res.* **61**, 385–389 (1973).
13. Goldberg, M. E. & Bushnell, M. C. *J. Neurophysiol.* **46**, 773–787 (1981).
14. Wurtz, R. H. & Mohler, C. W. *J. Neurophysiol.* **39**, 766–772 (1976).
15. Moran, J. & Desimone, R. *Science* **229**, 782–784 (1985).
16. Chelazzi, L., Miller, E. K., Duncan, J. & Desimone, R. *Nature* **363**, 345–347 (1993).
17. Motter, B. C. *J. Neurosci.* **14**, 2178–2189 (1994).
18. Lamme, V. A. F. *J. Neurosci.* **15**, 1605–1615 (1995).
19. Maunsell, J. H. R. *Science* **270**, 764–769 (1995).

20. Thompson, K. G., Hanes, D. P. & Schall, J. D. *Soc. Neurosci. Abstr.* **21**, 1270 (1995).
21. Maunsell, J. H. R. & Gibson, J. R. *J. Neurophysiol.* **68**, 1332–1344 (1992).
22. Nowak, L. G., Munk, M. H. J., Girard, P. & Bullier, J. *Vis. Neurosci.* **12**, 371–384 (1995).
23. Weinberger, N. M. A. *Rev. Neurosci.* **18**, 129–158 (1995).
24. Hanes, D. P., Thompson, K. G. & Schall, J. D. *Expl Brain Res.* **103**, 85–96 (1995).

ACKNOWLEDGEMENTS. Authorship is alphabetical. We thank D. Hanes for data, and R. Blake, F. Ebner, J. Kaas and D. Hanes for comments on the manuscript. This work was supported by the National Eye Institute and the McDonnell–Pew Program in Cognitive Neuroscience. J. Schall is a Kennedy Center investigator.

CORRESPONDENCE and requests for materials should be addressed to J.D.S. (e-mail: schalljd@ctrax.vanderbilt.edu).

The neurobiology of sign language and its implications for the neural basis of language

Gregory Hickok*, Ursula Bellugi* & Edward S. Klima*†

* The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

† The University of California, San Diego, California 92186, USA

The left cerebral hemisphere is dominant for language, and many aspects of language use are more impaired by damage to the left than the right hemisphere. The basis for this asymmetry, however, is a matter of debate; the left hemisphere may be specialized for processing linguistic information¹⁻³ or for some more general function on which language depends, such as the processing of rapidly changing temporal information⁴ or execution of complex motor patterns⁵. To investigate these possibilities, we examined the linguistic abilities of 23 sign-language users with unilateral brain lesions. Despite the fact that sign language relies on visuo-spatial rather than rapid temporal information, the same left-hemispheric dominance emerged. Correlation analyses of the production of sign language versus non-linguistic hand gestures suggest that these processes are largely independent. Our findings support the view that the left-hemispheric dominance for language is not reducible solely to more general sensory or motor processes.

Like spoken languages, sign languages used by the deaf are highly structured linguistic systems, with a rigid developmental course, including a critical period for acquisition⁶. There is no universal sign language, nor are they manual forms of surrounding spoken languages: American sign language (ASL) and British sign language, for example, are mutually incomprehensible. Signed languages have linguistic structure at phonological, morphological and syntactic levels. At the phonological level, signs are fractionated into sublexical elements, including recurring hand shapes, articulation locations, and limb/hand movements^{7,8}. There are even systematic 'phonetic' differences between sign languages leading to an 'accent' when native users of one sign language learn another^{3,9}. At the morphological level, ASL has developed grammatical markers that serve as inflectional and derivational morphemes⁹. At the syntactic level, ASL specifies relations between signs through, among other things, the manipulation of signs in space, where different spatial relations convey systematic differences in meaning¹⁰⁻¹² (Fig. 1).

Thus, although sign language has linguistic structuring at the same levels as spoken language, the surface form is radically different, with spatial contrasts prominent at every level. The time course between the shortest linguistically relevant transitions during a sign, such as a change in hand shape, is approximately 200 ms (ref. 13), significantly longer than the time course of the shortest transitions within a spoken word (~40 ms) that have been

TABLE 1 Biographical and medical data of subjects

	Age at sign exposure (years)	Age at onset of deafness (years)	Sex	Handedness	Age at testing (years)	Lesion size/location	Lesion aetiology
Left-lesioned							
LHD01	6	5	M	R	81	lg/front-par	Ischaemic infarct
LHD02	5	5	F	R	66	mod/inf par	Ischaemic infarct
LHD03	0	0	F	R	37	lg/front	Ischaemic infarct
LHD04	6	1	F	R	51	sm/inf-ant front	Aneurism rupture*
LHD05	13	0	M	R	45	lg/temp-par	Haematoma
LHD06	0	0	M	R	77	mod/front-temp-par	Ischaemic infarct
LHD07	0	0	M	R	86	sm/sup front-par	Ischaemic infarct
LHD08	6	2	F	R	64	mod/medial occ	Ischaemic infarct
LHD09	7	<1	M	R	29	mod/front-par	Ischaemic infarct
LHD10	0	2	F	R	79	mod/inf-post front	Ischaemic infarct
LHD11	9	<1	F	R	73	mod/front-par	Ischaemic infarct
LHD12	11	0	F	R	79	lg/front-temp-par	Ischaemic infarct
LHD13	4	0	M	R	71	mod/inf front-par	Haematoma
Right-lesioned							
RHD01	12	0	F	R	71	lg/front-temp-par	Ischaemic infarct
RHD02	9	5	M	R	82	mod/temp-par	Ischaemic infarct
RHD03	5	0	M	R	60	lg/front-temp-par	Ischaemic infarct
RHD04	0	0	F	R	61	mod/sup front-par	Tumour*
RHD05	0	n/a	F	R	38	mod/sup par-occ	Haematoma*
RHD06	0	0	M	R	74	lg/front-temp-par	Ischaemic infarct
RHD07	11	2	F	R	78	mod/front-par	Ischaemic infarct
RHD08	7	<1	M	R	74	lg/front-temp-par	Ischaemic infarct
RHD09	6	3	F	R	83	mod/temp-par	Ischaemic infarct
RHD10	0	0	F	R	78	mod/temp-par-occ	Ischaemic infarct

Ages are rounded to the nearest year. Lesions are grouped into three size categories based on visual inspection: lg, large; mod, moderate; sm, small. Lesion location is indicated by the lobe(s) involved: front, frontal; temp, temporal; par, parietal; occ, occipital; and by position within the lobe: sup, superior; inf, inferior; ant, anterior; post, posterior.

* Surgical intervention.

argued to underlie left-hemisphere dominance for language. Determining the lateralization of signed language therefore bears directly on the issue of the role of fast temporal processing in hemispheric dominance for language.

There have been only a few case studies investigating the role of the left cerebral hemisphere in processing sign language^{3,14,15}, and no quantitative group-level analyses comparing sign-language users with damaged left and right hemispheres. Here we report on a relatively large group of 13 left-hemisphere-damaged (LHD) signers and 10 right-hemisphere-damaged (RHD) signers (Table 1).

Using an ASL-adapted version of the Boston Diagnostic Aphasia Examination¹⁶, we assessed each subject's competence in several aspects of language use: production, comprehension, naming and repetition. LHD signers performed significantly worse than RHD signers on all measures (Fig. 2a). The differences apply even if subjects with lesions outside the perisylvian region are excluded (subjects in analyses, 10 LHD, 7 RHD; for each measure, $P < 0.03$). Finally, the difference between LHD and RHD signers is not a function of sampling error due to group differences in age at test, age of onset of deafness, or age of exposure to ASL. Across the two groups, there is no correlation between the total score on Boston Diagnostic Aphasia Examination rating scales¹⁶ and these three variables ($P = 0.99$, 0.52 and 0.91, respectively).

To ensure that deficits in sign-language processing were not simply a function of deficits in general spatial cognitive ability, standard measures of visuo-spatial cognition were administered. Figure 2b presents examples of the performance on these visuo-spatial tasks for four aphasic LHD signers, who nonetheless performed well on these visuo-spatial

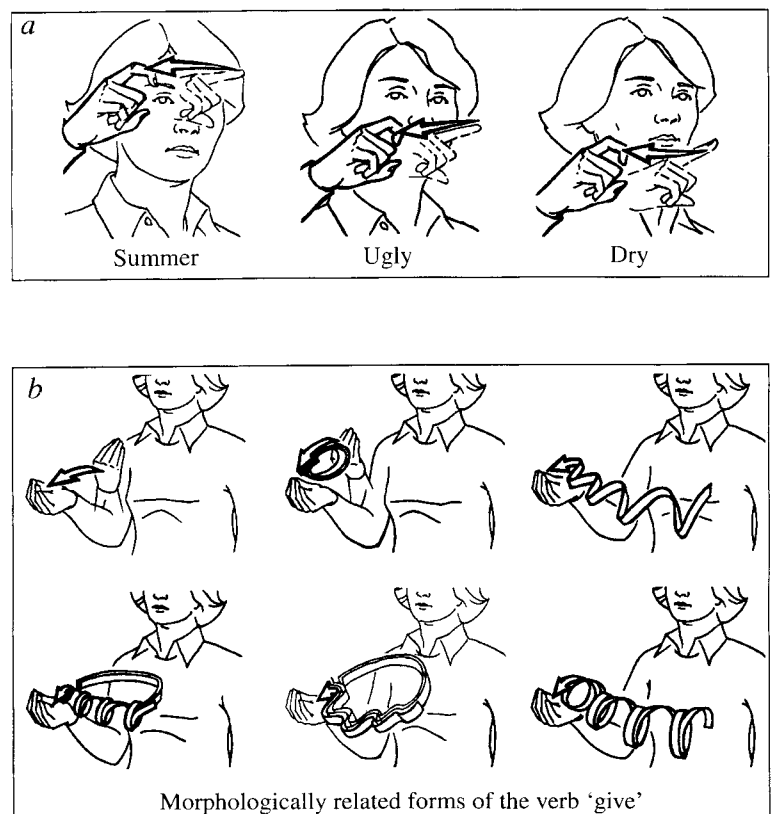


FIG. 1 a, Spatial contrasts at the lexical level. Articulation at different spatial locations relative to the body yields distinct signs. b, Spatial modulations in ASL morphology. Examples of how verb signs can be inflected through the modulation of movement, and how these inflections can be layered (embedded) within one another to derive subtle differences in meaning. Clockwise from top left: Give (uninflected); Give (continuous), 'give repeatedly'; Give (exhaustive), 'give to each in turn'; Give ((continuous) exhaustive), 'give repeatedly to each in turn'; Give (((continuous) exhaustive) continuous), 'give repeatedly to each in turn again and again'.

tasks, and four non-aphasic RHD signers, who performed poorly on these visuo-spatial tasks. This double dissociation between visuo-spatial abilities and sign language abilities indicates that these two domains of cognition are largely independent in deaf signers.

These data indicate that at the hemispheric level the neural organization of sign language is indistinguishable from that of spoken language. Given that sign language relies largely on spatial information rather than rapidly changing temporal information to encode linguistic distinctions, our findings demonstrate that

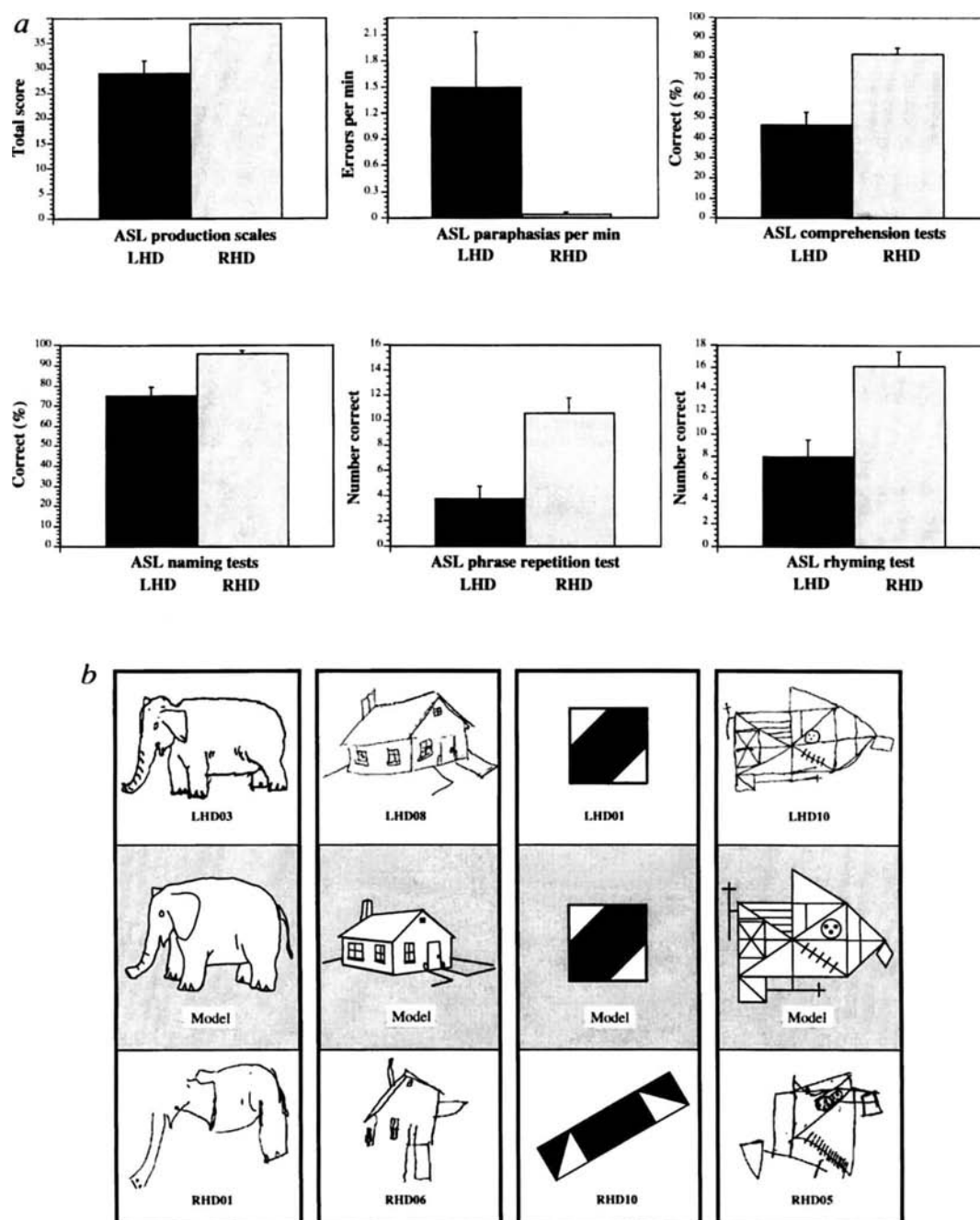


FIG. 2 *a*, Six measures of ASL ability for LHD versus RHD signers. All LHD signers and 9 of 10 RHD signers are deaf and use ASL as their primary means of communication; one RHD signer is a hearing ASL interpreter. Subjects with bilateral lesions or small lacunar infarcts were excluded. Subjects were tested at least three months after the stroke. Deaf, native ASL signers administered and scored all tests. Significance levels corrected for the number of tests, two-tailed: $P = 0.0002$. The production scales measure used the sum score on our ASL-adapted version of the Boston Diagnostic Aphasia Examination (BDAE) profile of speech characteristics¹⁶. Paraphasias per minute is the total number of sign errors in a sign sample elicited according to the BDAE protocol; $P = 0.02$ (excludes one LHD outlier, who produced errors at a rate 3 s.d. above the LHD mean). Comprehension tests were the BDAE commands subtest and our ASL-

adapted version of the token test¹⁸; two subjects (1 LHD, 1 RHD) did not take these tests; $P = 0.016$. Naming tests are the BDAE visual confrontation and responsive naming tests; $P = 0.017$. The phrase repetition test was an ASL version of the BDAE phrase repetition test; $P = 0.001$. The rhyming test was a 'rhyme' judgement test in which subjects choose (out of an array of four) the two pictured objects whose signs were most similar in terms of sign-phonological features³; $P = 0.009$; 10 LHD and 7 RHD subjects took this test. *b*, Examples of performance on visuo-spatial tasks by LHD signers (top) and RHD signers (bottom), along with the target stimulus (middle). Tests include the BDAE drawing to copy subtest¹⁶, the Wechsler adult intelligence scale-revised block design test, and the Rey Osterrieth complex figure¹⁹.

left-hemisphere dominance for language is not solely determined by a general proclivity for processing fast temporal information.

To investigate whether sign-language disruption can be reduced to disruptions of domain-general motor control³, we administered an abbreviated version³ of Kimura's movement copy test^{5,17} to 11 of the LHD subjects; they were asked to copy non-representational manual movements using the arm ipsilateral to the lesion. We found varying degrees of disruption in the ability to perform this task; however, scores did not correlate significantly with measures of sign production during connected sign, including number of paraphasias per minute ($r = 0.36$, $P = 0.27$), number of paraphasias when corrected for number of signs produced ($r = 0.32$, $P = 0.33$), or fluency as defined in the Boston

Diagnostic Aphasia Examination phrase-length scale ($r = 0.21$, $P = 0.54$). Further, on each of the language measures, subjects could be identified who produced similar scores in terms of their sign production yet differed substantially in their apraxia score (Fig. 3), indicating the dissociability between the two domains. Although it is difficult to rule out fully the existence of a significant correlation between these variables because of the relatively small sample size, these data suggest that there is a significant amount of variability in at least some aspects of sign language disruption that cannot be accounted for solely by a disruption of motor control.

Taken together, these data suggest that left-hemisphere dominance for language is not driven by physical characteristics of the linguistic signal or motor aspects of its production, but rather stem from higher-order properties of the system. Whether these turn out to be domain-specific aspects of grammatical structure or other, less specific, organizational properties awaits further investigation. □

Received 15 February; accepted 18 April 1996.

1. Bellugi, U., Poizner, H. & Klima, E. *Trends Neurosci.* **10**, 380–388 (1989).
2. Corina, D. P., Jyotsna, V. & Bellugi, U. *Science* **255**, 1258–1260 (1992).
3. Poizner, H., Klima, E. S. & Bellugi, U. *What the Hands Reveal About the Brain* (MIT Press, Cambridge, MA, 1987).
4. Tallal, P., Miller, S. & Fitch, R. H. *Ann. N.Y. Acad. Sci.* **682**, 27–47 (1993).
5. Kimura, D. *Neuromotor Mechanisms in Human Communication* (Oxford Univ. Press, 1993).
6. Newport, E. & Meier, R. in *The Crosslinguistic Study of Language Acquisition Vol. 1. The Data* (ed. Slobin, D. I.) (Lawrence Erlbaum, Hillsdale, NJ, 1985).
7. Corina, D. & Sandler, W. *Phonology* **10**, 165–207 (1993).
8. Perlmutter, D. M. *Linguistic Inquiry* **23**, 407–442 (1992).
9. Klima, E. & Bellugi, U. *The Signs of Language* (Harvard Univ. Press, Cambridge, MA, 1979).
10. Lillo-Martin, D. *Universal Grammar and American Sign Language: Setting the Null Argument Parameters* (Kluwer Academic, Boston, MA, 1991).
11. Lillo-Martin, D. & Klima, E. S. in *Theoretical Issues in Sign Language Research Vol. 1. Linguistics* (eds Fischer, S. D. & Siple, P.) 191–210 (Univ. Chicago Press, 1990).
12. Liddell, S. *American Sign Language Syntax* (Mouton, New York, 1980).
13. Corina, D. P. in *Phonetics and Phonology: Current Issues in ASL Phonology* (ed. Coulter, G. R.) 63–95 (Academic, New York, 1993).
14. Chiarello, C., Knight, R. & Mandel, M. *Brain* **105**, 29–51 (1982).
15. Poizner, H. & Kegi, J. *Ann. N.Y. Acad. Sci.* **682**, 192–213 (1993).
16. Goodglass, H. & Kaplan, E. *The Assessment of Aphasia and Related Disorders*. (Lea & Febiger, Philadelphia, 1976).
17. Kimura, D. *Phil. Trans. R. Soc. Lond. B* **298**, 135–149 (1982).
18. De Renzi, E. & Vignolo, L. A. *Brain* **85**, 665–678 (1962).
19. Osterrieth, P. A. *Archs. Psychol.* **30**, 206–356 (1944).

ACKNOWLEDGEMENTS. We thank C. Batch, K. Clark, M. Kritchewsky, K. Say and M. Withers for assistance, and the subjects and their families for participating in these studies.

CORRESPONDENCE and requests for materials should be addressed to G.H. (e-mail: hickok@crl.ucsd.edu).

Rapid and opposite effects of BDNF and NGF on the functional organization of the adult cortex *in vivo*

Neal Prakash, Susana Cohen-Cory*† & Ron D. Frostig

Department of Psychobiology and the Center for Learning and Memory, University of California at Irvine, Irvine, California 92717, USA

* Division of Biology, Caltech, Pasadena, California 91125, USA

THE adult cortex is thought to undergo plastic changes that are closely dependent on neuronal activity (reviewed in ref. 1), although it is not yet known what molecules are involved. Neurotrophins and their receptors have been implicated in several aspects of developmental plasticity^{2–4}, and their expression in the adult cortex suggests additional roles in adult plasticity^{5–9}. To examine these potential roles *in vivo*, we used

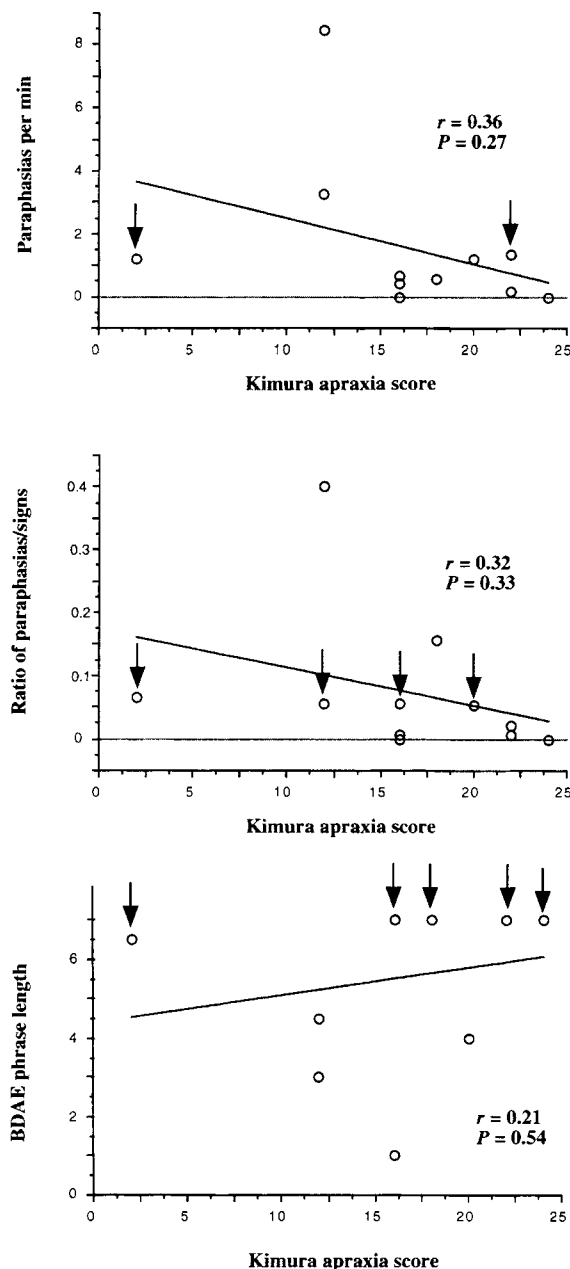


FIG. 3 Scatter plot and regression line for Kimura movement copy test score plotted against number of paraphasias (errors in sign) per minute of signing (top); ratio of paraphasias to total signs produced (middle); and BDAE phrase length scale (bottom). Note that significant variability in apraxia score does not necessarily predict variability in language measures (arrows).

† Present address: Mental Retardation Research Center, University of California Los Angeles, School of Medicine, Los Angeles, California 90024, USA.

intrinsic-signal optical imaging to quantify the effects of nerve growth factor (NGF) and brain-derived neurotrophic factor (BDNF) on the functional representation of a stimulated whisker in the 'barrel' subdivision of the rat somatosensory cortex. Topical application of BDNF resulted in a rapid and long-lasting decrease in the size of a whisker representation, and a decrease in the amplitude of the activity-dependent intrinsic signal. In contrast, NGF application resulted in a rapid but transient increase in the size of a representation, and an increase in the amplitude of the activity-dependent intrinsic signal. These results demonstrate that neurotrophins can rapidly modulate stimulus-dependent activity in adult cortex, and suggest a role for neurotrophins in regulating adult cortical plasticity.

In this study, we exploited the unique advantages of the barrel subdivision of the rat somatosensory system for the quantification of functional activity. Specifically, this system provides a one-to-one relationship between each whisker on the rat's snout and its discrete activity pattern in the contralateral barrel cortex. Moreover, this system is unique in providing the ability precisely to stimulate an individual whisker, which constitutes a single point on the sensory epithelium, enabling a straightforward interpretation of the evoked activity patterns in the cortex. This evoked sensory activity is coupled with characteristic spatiotemporal changes in the optical reflectance of the cortical tissue, termed 'intrinsic signals', that can be captured with a sensitive camera. By converting activity patterns into light reflectance patterns, intrinsic-signal optical imaging enables the visualization of a cortical functional representation (area of stimulus-dependent activation) with high spatial resolution ($\sim 50\ \mu\text{m}$)^{10–13}; changes in the size of a functional representation can therefore be quantified^{14,15}. Further, the amplitude of the stimulus-dependent intrinsic signal, which is directly correlated with the strength of stimulus-dependent neuronal activity¹⁶, can be quantified. This imaging system is

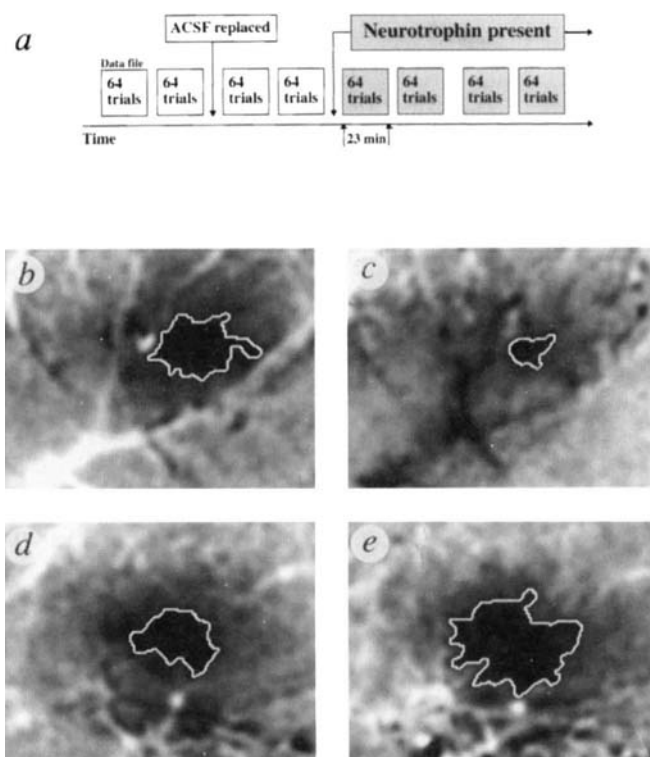
unique in allowing the visualization of events involved in cortical plasticity from large areas of the cortex *in vivo*.

Neurotrophins have been implicated in several aspects of developmental cortical plasticity^{2–4}, but little is known of their potential roles in adult plasticity. Here we examined the potential roles of BDNF and NGF on the modulation of stimulus-dependent cortical activity in the adult rat *in vivo*. The effects of neurotrophins on cortical activation were quantified over time, within individual anaesthetized rats, by comparing the area of functional representation evoked by the stimulated whisker before and after neurotrophin application (topical application of NGF or BDNF: $20\ \text{ng}\ \mu\text{l}^{-1}$ in artificial cerebrospinal fluid (ACSF)) (Fig. 1a). In addition, the amplitude of the response within the activated area was quantified by measuring the amplitude of the activity-dependent intrinsic signals before and after neurotrophin application. Topical application of BDNF resulted in a rapid and long-lasting decrease in the size of the functional representation of the stimulated whisker (Figs 1b, c and 2a), with a corresponding significant decrease in the amplitude of the intrinsic signal (Fig. 2c). With BDNF treatment, the size of the representation was decreased on average by 61% ($P = 0.04$; Fig. 2a), and the amplitude of the intrinsic signal was decreased on average by 21% ($P = 0.04$; Fig. 2c). Although the size of the functional representation was decreased by the treatment, the location of the centre of activity remained the same (Fig. 1b, c). The decrease in the size of the functional representation occurred rapidly, within 10 to 15 min of the topical application of BDNF; (Fig. 3a) this effect lasted for the duration of the experiment (up to 3 h with BDNF).

Immunostaining with an antibody for exogenous BDNF allowed visualization of BDNF penetration. Within 2 h of topical application, BDNF immunoreactivity was detected 150–180 μm from the cortical surface, suggesting that exogenous

FIG. 1 BDNF and NGF have opposite effects on the size of the functional representation of cortical activity following whisker stimulation. The cortical response to single-whisker stimulation was visualized and quantified by intrinsic-signal imaging before and after topical application of neurotrophin. **a**, Schematic representation of the protocol for each individual animal. The boxes represent eight data files, each consisting of 64 imaging trials. Four data files were collected with application of only ACSF (white boxes), and four with the application of recombinant human (rh)BDNF or rhNGF in ACSF (grey boxes). As an internal control to account for the effects of changing fluids over the exposed cortex, ACSF was replaced between the second and third data files. **b**, An example of the functional representation obtained by intrinsic-signal imaging following stimulation of whisker C2 on the contralateral snout of the rat before BDNF application. In this example, the white border delineates the functional representation (area at half-height; see Fig. 2 Methods) that equals $1.50\ \text{mm}^2$. **c**, With BDNF present, the functional representation decreased to $0.32\ \text{mm}^2$. **d**, Functional representation before NGF application; white border delineates the functional representation that equals $1.20\ \text{mm}^2$. **e**, With NGF present, the functional representation increased to $2.80\ \text{mm}^2$. All images are from single data files, and show an area of somatosensory cortex $6 \times 4.6\ \text{mm}$. The white and black streaks represent cortical blood vessels.

METHODS. See refs 13–15 for details. Adult male Sprague–Dawley (Charles River) rats (300–500 g) were initially anaesthetized with Nembutal (50 mg per kg; intraperitoneal). Throughout the rest of the experiment, Nembutal was injected as needed to maintain a deep anaesthetic state. After initial localization of the C2 whisker representation by imaging through the thinned skull, the skull and dura overlying the representation were removed. Next, a wall of petroleum jelly was built around the exposed cortex, filled with ACSF, and sealed with a coverglass. For neurotrophin treatment, we added BDNF or NGF to ACSF (each at $20\ \text{ng}\ \mu\text{l}^{-1}$), and the solution was left in the well for the remainder of the experiment (2 or 3 h). The cortex was illuminated with a 630-nm light and imaged using a slow-scan CCD camera (Photometrics). An imaging trial consisted of recording intrinsic-signal responses from the cortex 1 s before, during, and 2.5 s after stimulation of the contralateral C2 whisker (9 consecutive frames of 500 ms). Whisker stimulation was given with a computer-controlled stimulator at 5 Hz for 1 s; deflection was 0.5 mm rostrocaudally 15 mm from the snout. Data files were created from the summation of 64 consecutive imaging trials with an inter-trial interval of 10 s.



BDNF primarily affected cells in the upper layers of the cortex (Fig. 4a, b).

In contrast to BDNF, topical application of NGF resulted in a rapid but transient increase in the size of the functional representation (Figs 1d, e and 2b), with a corresponding increase in the

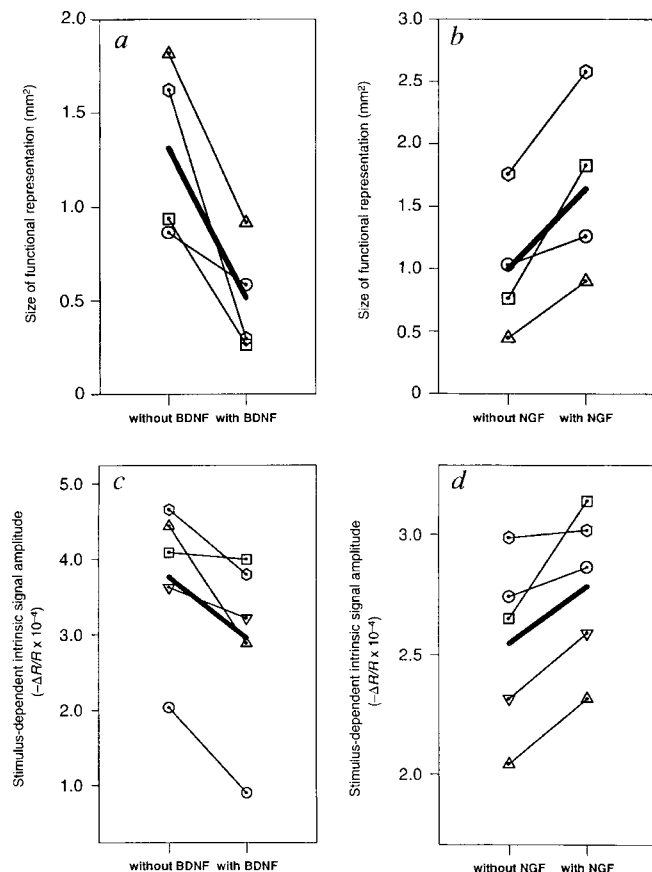


FIG. 2 BDNF (a, c) and NGF (b, d) have opposite effects on both the size of the functional representation and the amplitude of the activity-dependent intrinsic signal. Scatter diagrams of individual animals before and after application of neurotrophin show the changes in size of the functional representation (a, b) and amplitude of the stimulus-dependent intrinsic signal (c, d). The symbols connected by thin lines represent the average values for four data files before, and after, neurotrophin application, for each individual animal. The bold line in each scatter diagram represents the average for all experiments. **a**, Change in functional representation following BDNF application. BDNF decreased the functional representation on average by 61% (paired *t*-test, $n = 4$, $t = 3.64$, $P = 0.04$). **b**, Change in functional representation following NGF application. NGF increased the functional representation on average by 64% (paired *t*-test, $n = 4$, $t = 3.47$, $P = 0.04$). **c**, Change in the amplitude of the stimulus-dependent intrinsic signal following BDNF application. BDNF decreased the amplitude on average by 21% (paired *t*-test, $n = 5$, $t = 2.97$, $P = 0.04$). **d**, Change in the amplitude of the stimulus-dependent intrinsic signal following NGF application. NGF increased the amplitude on average by 9% (paired *t*-test, $n = 5$, $t = 3.71$, $P = 0.02$). All statistical testing was done using Sigma-Stat 2.0 (Jandel). For c and d, $\Delta R/R$ on the y-axis is the fractional change in reflectance.

METHODS. See refs 13–15 for details. A ratio file was obtained from each data file with a pixel-by-pixel division of the frames containing the raw reflectance data taken following whisker stimulation by the frames containing raw reflectance data taken from the pre- and post-response baseline data. The size of the functional representation was quantified for each ratio file by calculating the area at half-height (that is, the area inside the isogram that lies halfway between the peak value and the median value). The stimulus-dependent intrinsic signal amplitude was calculated from the raw reflectance data by preselecting a region of maximum overlap between all areas at half-height that did not overlie a major blood vessel.

amplitude of the intrinsic signal (Fig. 2d). During the first 2 h of NGF treatment, the size of the representation increased on average by 64% ($P = 0.04$; Fig. 2b), and the amplitude of the intrinsic signal increased on average by 9% ($P = 0.02$; Fig. 2d). Although the size of the cortical functional representation was increased by the treatment, the location of the centre of activity remained the same (Fig. 1d, e). The increase in the size of the functional representation occurred rapidly, starting within 10 to 15 min and peaking 30 min after topical application of NGF. Unlike the BDNF effect, the NGF effect was transient, returning to baseline over the next 90 min (Fig. 3b). In an independent set of

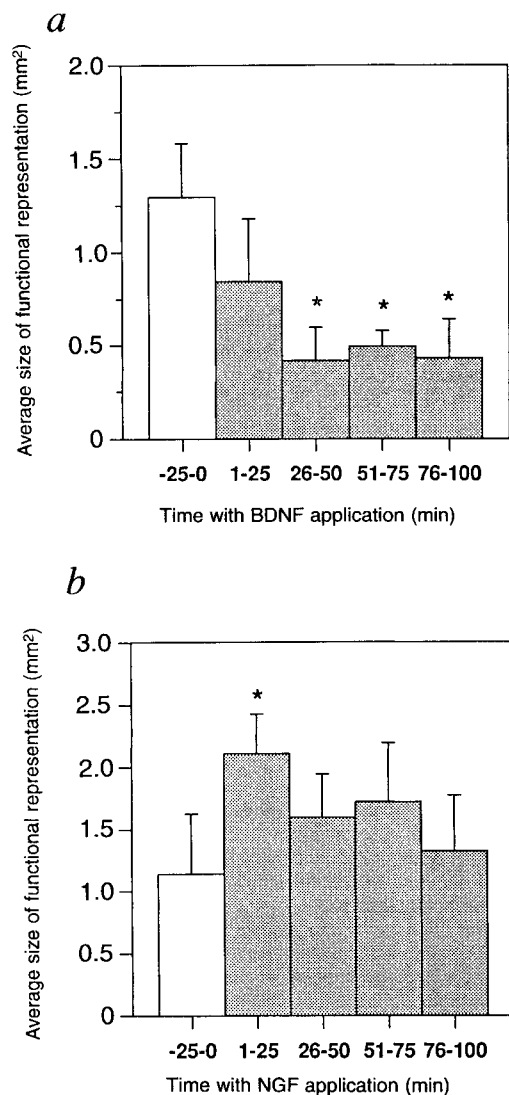


FIG. 3 Time course of **a**, BDNF, and **b**, NGF effects on the size of the functional representation. The white bar denotes the average size of the functional representation for the data files just before neurotrophin application. The grey bars are the average size of the functional representation at successive time bins in the continual presence of neurotrophin. Each data file was assigned to a time bin based on when the first trial of that data file was initiated relative to the time the neurotrophin was initially applied to the cortex. Error bars represent s.e.m.; asterisks denote statistical significance with $P < 0.05$. Statistical testing was with a one-way repeated-measure analysis of variance (ANOVA) and a post-hoc Dunnett's test. **a**, Rapid and long-lasting (sustained) effect of BDNF on the size of the functional representation. The effect reached and maintained significance after 26 min. Note the strong decrease within the first 25 min. **b**, Rapid but transient effect of NGF on the size of the functional representation. A significant effect is seen within the first 25 min. The trend remains above pretreatment level for the following 100 min.

experiments, the response to both neurotrophins was tested consecutively within the same rat. The NGF-elicited increase in the size and amplitude of the functional representation was reversed after a washout of excess NGF followed by BDNF application. This reversal was manifested as changes in the functional representation similar to those of BDNF application (that is, the average area and amplitude decreased; data not shown).

After 2 h of topical NGF application, exogenous NGF immunoreactivity was detected 200–250 μm from the cortical surface (Fig. 4c, d), indicating penetration to the supragranular layers. However, within 30 min of continuous NGF presence, NGF immunoreactivity was detected 80–100 μm from the cortical surface, suggesting that, at the peak of the NGF effect, NGF penetrated primarily to cortical layer I. The upper layers of the cortex (layers I–III) contain both neuronal cell bodies and projections from deeper cortical layers. Because intrinsic-signal imaging visualizes activity mostly from layers I–IV (ref. 17), it measures the summed activity both of cells located in these layers and of projections from deeper layers. This suggests that, although NGF and BDNF only penetrate the upper cortical layers, they can potentially affect neuronal populations at different cortical levels.

The localization of TrkA and TrkB, the high-affinity receptors for NGF and BDNF, respectively, in the adult rat barrel cortex is shown in Fig. 4e, f. Immunopositive cells for the full-length, functional form of the TrkB receptor localized mostly to the supragranular (Fig. 4e) and infragranular layers. TrkA-immunopositive fibres were seen to project to all cortical layers, including layer I (Fig. 4f). These findings are in agreement with previous work showing the expression of p75NTR⁷ (the low-affinity neurotrophin receptor), TrkB^{5,6} and *trkA*⁸ in the adult rat cortex, and suggest that cortical neurons have the receptors needed to respond to BDNF and NGF. Moreover, the localization of TrkA-immunopositive fibres that project intracortically or from other brain areas (such as basal forebrain)⁹ suggests an additional mechanism for NGF action.

The mechanisms by which BDNF and NGF exert their effects on the functional organization of the adult neocortex are not yet known. BDNF may affect multiple neuronal populations to produce the observed net effects, or may specifically enhance the activity of inhibitory neurons. In the central nervous system, BDNF has been associated with effects on multiple neuronal populations, such as GABA-ergic^{3,18,19} serotonergic¹⁹, dopaminergic¹⁹ and glutamatergic²⁰ neurons. In the adult neocortex, the BDNF receptor TrkB is expressed by a subpopulation of GABA-ergic cells²¹, suggesting that these inhibitory neurons respond to BDNF. The opposite effects of NGF suggest different mechanisms of action from those of BDNF. NGF has been associated with effects on cholinergic²² and glutamatergic²³ neurons in several areas in the adult brain. For instance, the cortex receives a projection from cholinergic neurons that respond to NGF and express the NGF receptor TrkA^{9,22}, suggesting that these neurons may mediate the cortical response to NGF. Therefore, differential actions of BDNF and NGF on different receptors expressed by distinct populations of neurons are the most plausible explanation for the opposite effects observed.

The involvement of neurotrophins in the modulation of synaptic efficacy *in vitro*, and the regulation of neurotrophin expression by neuronal activity, have

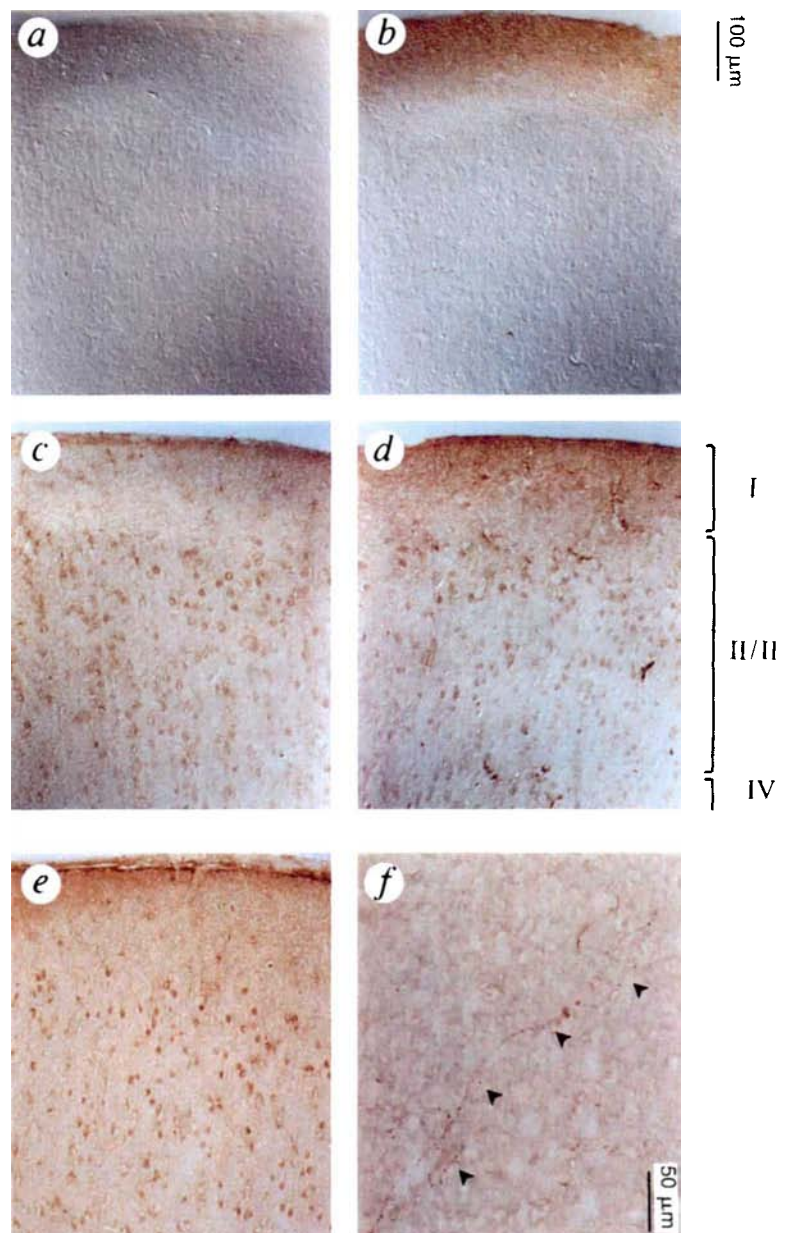


FIG. 4 Penetration of BDNF and NGF following topical application (a–d), and localization of TrkB and TrkA receptors (e, f), in the adult barrel cortex. a–d, Photomicrographs of cryostat sections from experimental brains treated with BDNF or NGF show immunostaining for rhBDNF or rhNGF in the adult barrel cortex. a, Antibody staining for exogenous BDNF showed only background staining on the untreated side. This particular antibody does not recognize the endogenously synthesized BDNF³⁰. b, On the treated side, BDNF immunoreactivity was detected at least 160–180 μm into the cortex. c, Antibody staining for NGF showed endogenous NGF-immunoreactive cells on the untreated side. d, On the treated side, NGF immunoreactivity was detected at least 200–220 μm from the cortical surface as seen by the diffuse darker staining of the neuropile, in addition to the NGF immunoreactive cells. e, f, Immunostaining with specific antibodies to TrkB or TrkA show the localization of TrkB receptors to neurons in the supragranular layers (layers I, II/III) of the barrel cortex, and TrkA-immunopositive fibres that project to these layers. e, The distribution of TrkB-positive cells is shown in a section adjacent to that shown in a and b (untreated side). f, The distribution of TrkA-immunopositive fibres is shown in a section adjacent to that shown in c and d (untreated side). Arrowheads point to a TrkA-labelled fibre that reached the upper part of layer II. The outline of the cortical layers was measured from cresyl-violet-stained sections adjacent to those shown in e and f. Scale bar for a–e, 100 μm ; for f, 50 μm .

METHODS. After each experiment, each rat was given an overdose of Nembutal and perfused with salinated phosphate buffer, and fixed with 4% paraformaldehyde. Brains were sectioned and processed for immunostaining using specific antibodies to rhBDNF³⁰, rhNGF (Santa Cruz), the intracellular domain of TrkB (TrkBIn; D. Kaplan) or the extracellular domain of TrkA⁹, followed by biotinylated secondary antibodies and visualization with the Vector ABC kit.

suggested reciprocal interactions between neurotrophins and neuronal activity. For example, experiments using adult barrel²⁴ and visual²⁵ cortices suggest that BDNF mRNA expression is regulated by neuronal activity. Conversely, BDNF also has a rapid effect on neuronal activity *in vitro*^{26–28}, enhancing synaptic transmission with a short time course^{26,27}. Although less is known about the interactions between NGF and activity-dependent plasticity in adult neocortical neurons, some evidence from other systems suggests a role for NGF in this process. For example, NGF is released in an activity-dependent manner in adult hippocampal slices²⁹, and can rapidly potentiate glutamate release and increase excitatory postsynaptic potentials in cultured hippocampal neurons²³. These findings suggest roles for BDNF and NGF in the modulation of neuronal activity, but direct evidence for their effects *in vivo* had been lacking. We have now demonstrated *in vivo* that BDNF rapidly decreases and NGF rapidly increases the cortical response to sensory stimulation. The effects of neurotrophins on the activity and functional organization of the adult cortex suggest a rapid neuromodulatory role for neurotrophins *in vivo*, along with their 'classical' roles as neurotrophic factors, and support a role for neurotrophins in adult cortical reorganization and plasticity. □

Received 3 January; accepted 28 April 1996.

1. Kaas, J. H. & Gilbert, C. D. in *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 51–90 (MIT Press, Cambridge, MA, 1995).
2. Maffei, L., Berardi, N., Domenici, L., Parisi, V. & Pizzorusso, T. *J. Neurosci.* **12**, 4651–4662 (1992).
3. Thoenen, H. *Science* **270**, 593–598 (1995).
4. Lindholm, D., Castrén, E., Berzaghi, M., Blöchl, A. & Thoenen, H. *J. Neurobiol.* **25**(11), 1362–1372 (1994).

5. Masana, Y., Wanaka, A., Kato, H., Asai, T. & Tohyama, M. *J. Neurosci. Res.* **35**, 468–479 (1993).
6. Zhou, X. F., Parada, L. F., Soppet, D. & Rush, R. A. *Brain Res.* **622**, 63–70 (1993).
7. Pitts, A. F. & Miller, M. W. *Somatosens. Mot. Res.* **12**, 329–342 (1995).
8. Valenzuela, D. M. et al. *Neuron* **10**, 963–974 (1993).
9. Holtzman, D. M. et al. *J. Neurosci.* **15**, 1567–1576 (1995).
10. Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D. & Wiesel, T. N. *Nature* **324**, 361–364 (1986).
11. Ts'o, D. Y., Frostig, R. D., Lieke, E. E. & Grinvald, A. *Science* **249**, 417–420 (1990).
12. Frostig, R. D., Lieke, E. E., Ts'o, D. Y. & Grinvald, A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6082–6086 (1990).
13. Masino, S. A., Kwon, M. C., Dory, Y. & Frostig, R. D. *Proc. natn. Acad. Sci. U.S.A.* **90**, 9998–10002 (1993).
14. Chen, C., Kwon, M. C., Masino, S. A. & Frostig, R. D. *J. Neurosci. Meth.* (in the press).
15. Masino, S. A. & Frostig, R. D. *Proc. natn. Acad. Sci. U.S.A.* **93**, 5022–5027 (1996).
16. Frostig, R. D., Masino, S. A. & Kwon, M. C. *Soc. Neurosci. Abstr.* **20**, 1384 (1994).
17. Malonek, D., Shoham, D., Ratzlaff, E. & Grinvald, A. *Soc. Neurosci. Abstr.* **16**, 292 (1990).
18. Mizuno, K., Carnahan, J. & Nawa, H. *Dev. Biol.* **165**, 243–256 (1994).
19. Altar, C. A. et al. *Exp. Neurol.* **130**, 31–40 (1994).
20. Leßmann, V., Gottmann, K. & Heumann, R. *NeuroReport* **6**(1), 21–25 (1994).
21. Cellerino, A., Burkhalter, A., Maffei, L. & Domenici, L. *Soc. Neurosci. Abstr.* **21**, 2010 (1995).
22. Cuello, A. C., Maysinger, D. & Garofalo, L. *Molec. Neurobiol.* **6**(4), 451–461 (1993).
23. Knipper, M., Leung, L. S., Zhao, D. & Rylett, J. *NeuroReport* **5**, 2433–2436 (1994).
24. Racamora, N., Welker, E., Soriano, E., Planas, A. M. & Van der Loos, H. *Soc. Neurosci. Abstr.* **19**, 259 (1993).
25. Castrén, E., Zafra, F., Thoenen, H. & Lindholm, D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 9444–9448 (1992).
26. Levine, E. S., Dreyfus, C. F., Black, I. B. & Plummer, M. R. *Proc. natn. Acad. Sci. U.S.A.* **92**, 8074–8077 (1995).
27. Kang, H. & Schuman, E. M. *Science* **267**, 1658–1662 (1995).
28. Knipper, M., Beck, A., Rylett, J. & Breer, H. *NeuroReport* **4**, 483–486 (1993).
29. Blöchl, A. & Thoenen, H. *Eur. J. Neurosci.* **7**, 1220–1228 (1995).
30. Nawa, H., Pelkeymounter, M. A. & Carnahan, J. *J. Neurosci.* **14**(6), 3751–3765 (1994).

ACKNOWLEDGEMENTS. We thank Amgen Inc. for BDNF; Genentech Inc. for NGF; J. Carnahan for BDNF antibodies; D. O. Clary for TrkB antibodies; D. Kaplan for TrkB antibodies; and S. E. Fraser for support and comments on the manuscript. This work was supported in part by NIMH, NSF, the UCI-Markey Program in Human Neurobiology (R.D.F.), the UCI Medical Scientist Training Program (N.P.), and the Helen Hay Whitney Foundation and NIMH (S.C.-C.).

CORRESPONDENCE and requests for materials should be addressed to R.D.F. (e-mail: rfrostig@uci.edu).

Regulation of synaptic responses to high-frequency stimulation and LTP by neurotrophins in the hippocampus

Alexander Figurov, Lucas D. Pozzo-Miller*, Petur Olafsson, Ti Wang & Bai Lu

Roche Institute of Molecular Biology, Nutley, New Jersey 07110, and Laboratory of Developmental Neurobiology, NICHD, National Institutes of Health, Bethesda, Maryland 20892, USA

NEUROTROPHINS promote neuronal survival and differentiation, but the fact that their expression is modified by neuronal activity, suggests a role in regulating synapse development and plasticity^{1–3}. In developing hippocampus, the expression of brain-derived neurotrophic factor (BDNF) and its receptor TrkB^{4–7} increases in parallel with the ability to undergo long-term potentiation (LTP)^{8–10}. Here we report a mechanism by which BDNF modulates hippocampal LTP. Exogenous BDNF promoted the induction of LTP by tetanic stimulation in young (postnatal day 12–13) hippocampal slices, which in the absence of BDNF show only short-term potentiation (STP). This effect was due to an enhanced ability of hippocampal synapses to respond to tetanic stimulation, rather than to a direct modulation of the LTP-triggering mechanism. A TrkB-IgG fusion protein, which scavenges endogenous BDNF¹¹, reduced the synaptic responses to tetanus as well as the magnitude of LTP in adult hippocampus.

Our results suggest that BDNF may regulate LTP in developing and adult hippocampus by enhancing synaptic responses to tetanic stimulation.

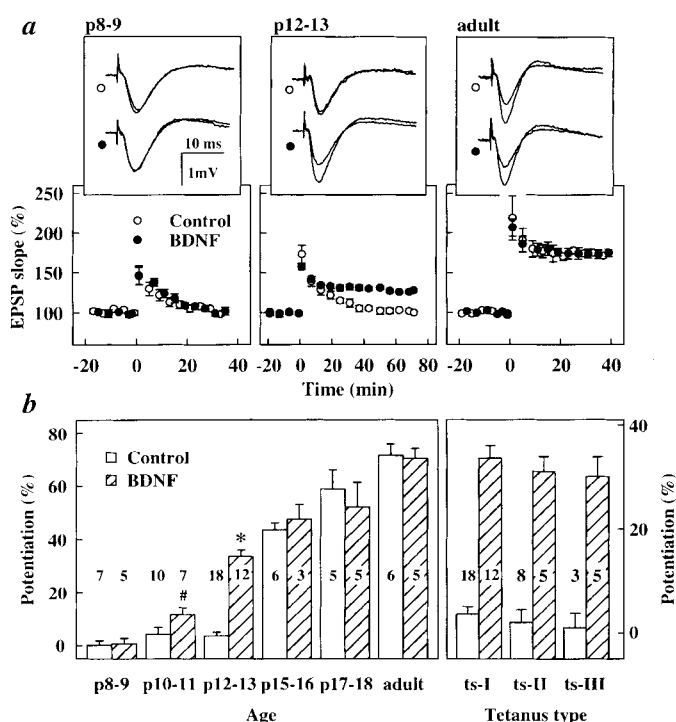
Field excitatory postsynaptic potentials (EPSPs) were elicited in hippocampal slices of different ages by stimulating two independent Schaffer–commissural afferents converging on the same CA1 neurons¹². Similar to previous observations^{8–10}, the magnitude and longevity of theta-burst stimulation (TBS)-induced synaptic potentiation increased during postnatal life (Fig. 1). In postnatal day 12–13 (p12–13) or younger slices, TBS elicited only STP that returned to baseline within 30–40 min (Fig. 1a). The first reliable TBS-induced LTP (LTP_T) occurred in p15–16 hippocampus, and its magnitude at p17–18 was similar to that in adult slices (Fig. 1b, left). In p12–13 slices (the subthreshold age group), TBS induced a typical STP lasting less than 50 min. In contrast, a stable LTP_T was observed when TBS was applied to a second pathway of the same slices after perfusion with BDNF (2 nM) (Fig. 1a); approximately 2.5 h of BDNF treatment was required to see its effects. Several different types of LTP-inducing tetanic stimulation (ts-II, 2 × 100 pulses at 100 Hz, 20 s apart; ts-III, 3 × 30 pulses at 100 Hz, 5 s apart) produced the same STP as TBS (ts-I) in control, and LTP_T after BDNF treatment (Fig. 1b, right). Thus the effect of BDNF on LTP_T was independent of tetanus parameters. In contrast to a previous report¹³, BDNF alone did not affect basal synaptic transmission elicited by low-frequency stimulation (LFS, 1 per min) at any ages examined under our experimental conditions. EPSP slope remained the same in p12–13 slices up to 3 h after treatment with BDNF (103 ± 3% of a baseline calculated 10 min before BDNF application; *n* = 11). Therefore, a combination of tetanic stimulation and BDNF was needed to elicit LTP_T in p12–13 slices.

The effect of BDNF on LTP_T was most prominent in a period before the onset of LTP_T. BDNF had no effect on synaptic potentiation in p8–9 slices (Fig. 1a). In p17–18 and adult hippocampus, standard LTP_T was observed in all slices (Fig. 1b, left). Exogenous BDNF no longer increased the magnitude of LTP_T, possibly because of high levels of endogenous BDNF^{4,5}. The

* Present address: Laboratory of Neurobiology, NINDS, National Institutes of Health, Bethesda, Maryland 20892, USA

FIG. 1 Effect of BDNF on LTP_I development. BDNF treatment: 2 nM for 2.5–4 h. **a**, TBS-induced potentiation of synaptic efficacy in CA1 region of hippocampal slices in p8–9, p12–13 and adult rats. Top, sample traces of field EPSPs recorded before and after TBS. Superimposed traces were taken 10 min before and 40–60 min after TBS. Bottom, time course of synaptic potentiation. Numbers of experiments are as indicated in the columns for the corresponding age groups in Fig. 1b, left. Synaptic efficacy (initial slope of field EPSPs) is expressed as the percentage of baseline value recorded during the 10 min before TBS. TBS was applied at time 'zero'. **b**, Left, summary of the BDNF effect on LTP_I at different ages as indicated. Percentages of potentiation (above baseline values) measured 40 min after TBS are calculated. A two-factor analysis of variance (ANOVA) and subsequent multiple tests were used to determine a significant effect of BDNF across age. BDNF increased synaptic efficacy over control at p10–11 (#, $P < 0.05$) and p12–13 (*, $P < 0.001$), but not in other age groups. Right, potentiation of synaptic efficacy induced by different tetanic stimulation in control and BDNF-treated p12–13 slices. All data were calculated as percentages of potentiation (above baseline values) measured 40 min after tetanic stimulation. Stimulation parameters: ts-I, TBS, 10 bursts, each of 4 pulses at 100 Hz; interburst interval, 200 ms; ts-II, 2 bursts of 100 pulses at 100 Hz, 20 s apart; ts-III, 3 bursts of 30 pulses at 100 Hz, 5 s apart. In all cases, tetanic stimulation induced STP in control, but LTP_I after BDNF treatment. There were no significant differences in the magnitudes of potentiation induced by different types of tetanic stimulation ($P > 0.1$, *t*-test) in either control and BDNF-treated slices.

METHODS. Hippocampal slices (400 μ m) were prepared¹² from male Sprague–Dawley rats (Harlan) of indicated ages and allowed to recover in an interface chamber for 1.5 h in artificial atmosphere of 95% O₂/5% CO₂. The slices were continuously perfused with artificial cerebrospinal fluid (ACSF) (34 °C) containing (in mM): NaCl, 124; KCl, 3.0; CaCl₂, 2.5; MgCl₂, 1.5; NaHCO₃, 26; KH₂PO₄, 1.25; glucose, 10; and ascorbate, 2, pH 7.4. Human recombinant BDNF or other neurotrophins (provided by Genentech, San Francisco) was added to ACSF at a final concentration of 2 nM. Field EPSPs were recorded extracellularly in CA1 area with glass microelectrodes filled with ACSF (resistance 4–8 M Ω). Monophasic stimuli (200 μ s duration) were delivered alternately (1 per min) by two stimulating electrodes positioned on either side of the recording microelectrode. Only slices exhibiting EPSPs of 2–3 mV amplitude without superimposed population spike were used. The stimulus strength was adjusted to evoke field EPSPs of approximately 1.3 mV. Tetanic stimulation was applied after a



period of control recording. In most cases, tetanic stimulation was applied to one pathway and EPSPs were monitored for at least 40 min. The same slices were then treated with the neurotrophins for 2.5–4 h, and tetanic stimulation was applied to the second pathway. All responses were digitized (10 kHz), analysed on-line and stored on magnetic media using acquisition systems DaQSys 1.0 and Process 2.0 (University of Geneva). Error bars in all figures are s.e.m.

expression of catalytic trkB mRNA also increases during the period from P6 to P13 in both CA3 and CA1 neurons⁷.

P12–13 slices incubated in artificial cerebrospinal fluid for 3.5 h did not exhibit LTP_I (Fig. 2), suggesting that a simple prolonged incubation *in vitro* does not lead to LTP_I. Nerve growth factor (NGF) and neurotrophin-3 (NT-3) which activate TrkA and TrkB receptor tyrosine kinases¹⁴, respectively, failed to have any effects on LTP_I development. In contrast, neurotrophin 4/5 (NT-4/5), which acts on the same TrkB receptor¹⁵, had an effect similar to BDNF. The BDNF effect was prevented by pretreatment of the slices for 20–30 min with a TrkB–IgG fusion protein (minimal effective concentration, 1 μ g ml⁻¹), a specific scavenger for BDNF¹¹. TrkB–IgG or IgG alone affected neither basal synaptic transmission nor LTP_I in p12–13 rats.

Tetanic stimulation elicits a rise in intracellular Ca²⁺ through N-methyl-D-aspartate receptor (NMDA) and voltage-dependent Ca²⁺ channels, leading to LTP induction^{16,17}. Application of the NMDA antagonist 2-amino-5-phosphonovaleric acid (AP5, 50 μ M) completely blocked LTP_I induction in BDNF-treated, p12–13 slices ($n = 7$), suggesting that the BDNF-facilitated LTP_I was NMDA dependent. However, BDNF did not affect NMDA-mediated EPSPs elicited by LFS (1 per min). EPSPs in the presence of the non-NMDA receptor antagonist 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX; 10 μ M) and low magnesium concentration (0.2 mM) remained unchanged up to 3 h after BDNF treatment ($101 \pm 2\%$ of a baseline calculated 10 min before BDNF application, $n = 4$). In p12–13 slices treated with bicuculline (3–10 μ M), a GABA_A receptor antagonist, TBS induced the same STP in control ($100 \pm 2\%$, $n = 6$) and LTP_I after BDNF treatment ($129 \pm 3\%$, $n = 10$), suggesting that the BDNF effect on LTP_I was not due to a decreased GABAergic inhibition.

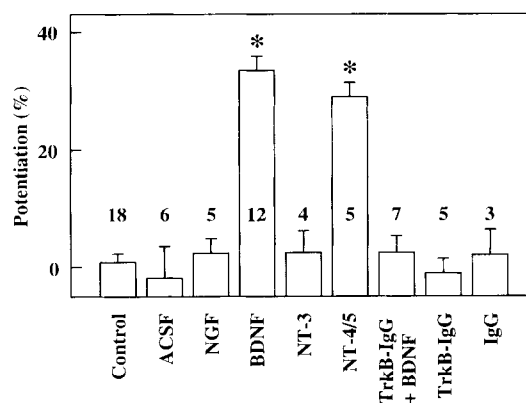


FIG. 2 Pharmacological characterization of neurotrophin effects on TBS-induced potentiation of synaptic efficacy in p12–13 hippocampal slices. Slices were treated with various neurotrophins (2 nM), TrkB–IgG fusion protein (1 μ g ml⁻¹), human IgG (1 μ g ml⁻¹), or ACSF for 2.5–4 h. Percentages of potentiation above baseline values were measured 40 min after TBS. 'Control' is synaptic efficacy measured 40 min after TBS before treatment of any of the above agents. TBS delivered to p12–13 slices after 2.5–4 h of incubation in ACSF did not exhibit LTP. In the 'TrkB–IgG + BDNF' group, slices were treated with TrkB–IgG for 20–30 min before application of BDNF. The numbers on the columns represent the total number of experiments performed. *, 'BDNF' and 'NT-4/5' groups are significantly higher than all other groups (one factor ANOVA and post-hoc comparison; $P < 0.001$).

Another possibility for LTP to occur in p12–13 hippocampus is that BDNF enhanced synaptic responses to the LTP-inducing tetanic stimulation, such as TBS. We compared EPSPs elicited by TBS (10 bursts, each with 4 pulses at 100 Hz; interburst interval, 200 ms) in control and BDNF-treated (2.5–4 h) slices. The size of EPSPs elicited by the second, third and fourth pulses in the first burst (1st-2, 1st-3 and 1st-4 in Fig. 3a) became much greater in BDNF-treated slices. Moreover, the synaptic responses to each of the burst stimuli gradually declined in control, but remained

increased in BDNF-treated slices. The greatest difference was observed in the response to the 4th pulse in the 10th burst (10th-4 in Fig. 3b). The effects of BDNF on synaptic responses to TBS were reminiscent of those on LTP_T development. The ability of hippocampal synapses to follow repetitive stimulation gradually increased with age (Fig. 3c). In BDNF-treated slices, the developmental profile of 1st-4 EPSP shifted to the 'left', and the most prominent effect of BDNF was observed in p12–13 slices (Fig. 3c). A minimum of 2.5 h BDNF treatment was required to see this effect (data not shown). The synaptic response to 1st-4 pulse of TBS was significantly reduced in p17–18 and adult slices treated with TrkB-IgG, but slices from younger animals were not affected (Fig. 3c), suggesting that endogenous BDNF is involved in modulating synaptic responses to tetanic stimulation. We also examined synaptic responses induced by 100 pulses at 100 Hz, another commonly used LTP-inducing tetanus. BDNF again attenuated the gradual decline of synaptic responses reflecting synaptic fatigue¹⁸ frequently seen in immature synapses (Fig. 3d, left). For example, a comparison of the 20th responses showed that the EPSP slope in BDNF-treated slices was 58% higher than that in control (Fig. 3d, right).

If a certain threshold of TrkB activation must be achieved for TBS to elicit LTP, we would predict that preventing TrkB activation should preclude LTP_T in adult hippocampus, where BDNF and trkB are highly expressed^{4–7}. Indeed, pretreatment with TrkB-IgG, which reduced synaptic responses to TBS (Fig. 3c), greatly attenuated LTP_T in adult slices (Fig. 4a). Furthermore, exogenous BDNF appeared to enhance the potency of tetanic stimulation. In adult control slices, weak tetanus (4 instead of 10 bursts of TBS) induced only an STP that lasted no longer than 40 min. The same weak tetanus elicited a sustained LTP_T in slices treated with BDNF (Fig. 4b). The LTP-triggering mechanism involves an activation of the NMDAR by either tetanus or post-synaptic depolarization, leading to a rise of intracellular Ca²⁺ and subsequent biochemical steps activated by this rise in Ca²⁺ (refs 16, 17). To determine whether BDNF directly affects LTP-triggering events downstream of NMDAR activation, we induced LTP by pairing low-frequency stimulation (0.05 Hz) with postsynaptic depolarization elicited by the K⁺-channel blocker tetraethylammonium (TEA, 25 mM), thereby avoiding tetanic afferent stimulation^{19–21}. In all p12–13 slices, pairing of presynaptic low-frequency stimulation with bath application of TEA elicited robust LTP (LTP_s), and BDNF treatment did not produce further potentiation (Fig. 4c, left). Moreover, TrkB-IgG did not block LTP_s in adult slices (Fig. 4c, right). Thus failure of LTP_T in young slices may reflect the inability of these synapses to follow repetitive stimulation, and BDNF influences LTP_T by enhancing the ability of hippocampal synapses to follow high-frequency stimulation, rather than by modulating the LTP-triggering mechanisms.

Neurotrophins can regulate neuronal activity and synaptic efficacy^{22–27}, but the physiological significance of this remains unclear. The fact that BDNF mRNA expression in the hippocampus is upregulated by LTP-inducing tetanic stimulation^{1,2} suggests a reciprocal interaction between neurotrophin expression and synaptic activity. Here we found that BDNF promotes LTP_T development in the hippocampus, supporting the hypothesis of a feedback loop. Moreover, BDNF modulates LTP_T by enhancing synaptic responses to high-frequency stimulation. Synaptic fatigue elicited by high-frequency stimulation is due to a depletion of releasable synaptic vesicles¹⁸. Attenuation of synaptic fatigue by BDNF not only implies a presynaptic effect, but also suggests a significant role of BDNF in modulating high-frequency synaptic activity *in vivo*. Consistent with BDNF knockout studies²⁸, a critical level of BDNF/TrkB activity appears to be important for long-term modulation of synaptic efficacy. Exogenous BDNF promoted synaptic responses to tetanus and enhanced LTP_T in neonatal hippocampus where endogenous BDNF level is low, whereas TrkB-IgG had the opposite effect in adult hippocampus, where endogenous BDNF level is high. BDNF/TrkB-mediated facilitation of activity- and NMDA-dependent synaptic potentia-

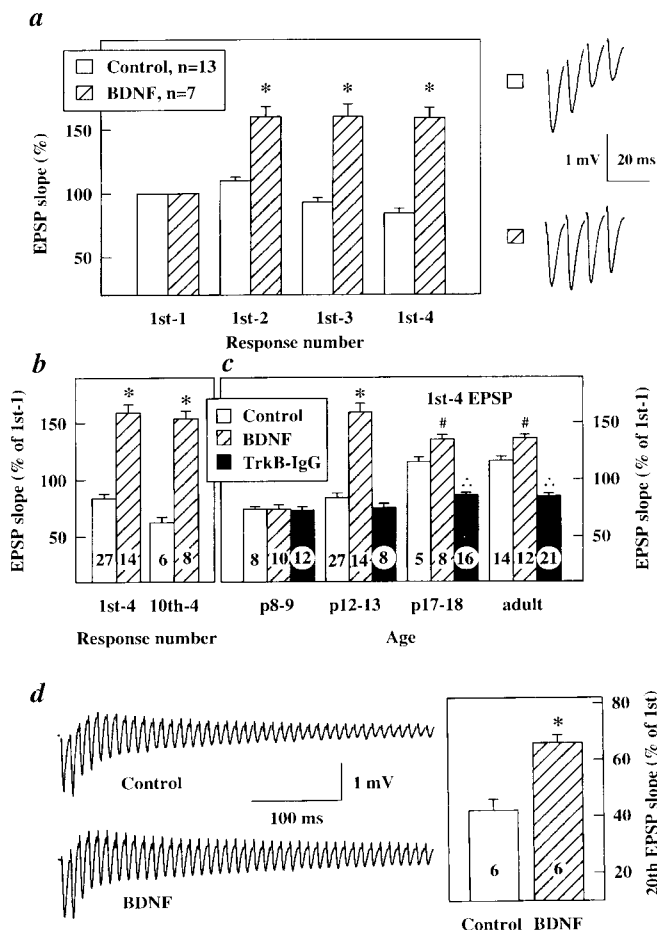


FIG. 3 Effects of BDNF and TrkB-IgG on synaptic responses to tetanic stimulation. BDNF treatment: 3–3.5 h at 2 nM. Number of experiments is indicated in each column. Student *t*-test was used in all the data except in c, in which a two-factor ANOVA and subsequent multiple tests were used; * and #, significantly higher than control at $P < 0.001$ and $P < 0.05$, respectively; . indicates significantly lower than control at $P < 0.001$. **a**, Synaptic responses to the first burst of TBS (four pulses at 10 ms interval) in p12–13 slices. The initial slope of field EPSPs was normalized to that of the first of the four responses. Traces of EPSPs in control and BDNF-treated slices are shown on the right. **b**, Attenuation of synaptic fatigue by BDNF in p12–13 slices. Synaptic responses to the last pulse in the first burst (1st-4) and that in the 10th burst (10th-4) of TBS in control and BDNF-treated slices are compared. All data are normalized to the first EPSP of the first burst (percentage of 1st-1). The EPSPs elicited by the last pulse of each burst progressively decreased in control but remained high in BDNF-treated slices. **c**, Effect of BDNF and TrkB-IgG on TBS-induced EPSPs during development. Only 1st-4 EPSPs are presented. There was a gradual increase in the synaptic response, and BDNF shifted the developmental profile to the left. TrkB-IgG ($1 \mu\text{g ml}^{-1}$, 2.5–3.5 h), in contrast, reduced the synaptic responses in p17–18 and adult slices. **d**, EPSPs induced by a long train of tetanus (100 Hz) in p12–13 slices. Left, partial display of field EPSPs induced by the tetanus in control and BDNF-treated slices. The decrease in the size of EPSPs is slower in BDNF-treated slices. Right, initial slope of the 20th EPSP (normalized to that of the first EPSP) in control and BDNF-treated slices.

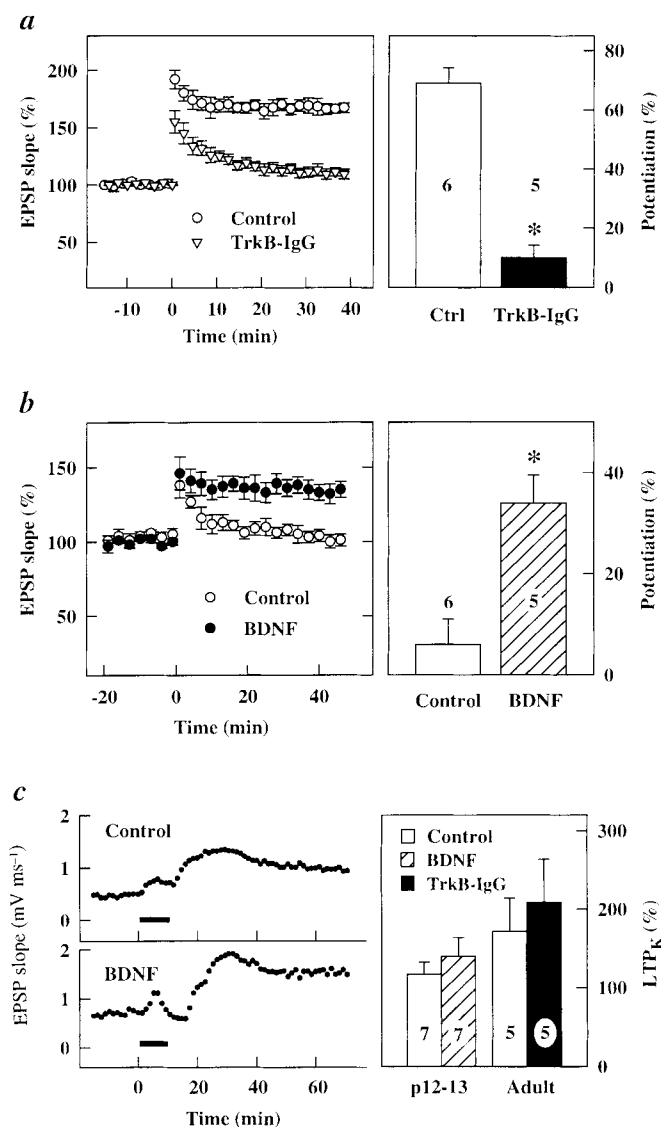


FIG. 4 Effects of BDNF and TrkB-IgG on LTP_T and LTP_K. The numbers associated with the columns represent the total number of experiments performed. *, *t*-test, *P* < 0.001. Left, time courses of synaptic potentiation; right, summaries of synaptic potentiation (per cent above baseline values). **a**, Effect of TrkB-IgG fusion protein on LTP_T. Note that LTP_T measured 40 min after TBS (right) was substantially attenuated in TrkB-IgG treated slices (1 µg ml⁻¹, 2.5–3.5 h). Control slices incubated in ACSF for similar duration exhibited normal LTP_T. **b**, Effect of BDNF on a weak tetanus-induced potentiation. A weak tetanic stimulation (4 bursts of TBS instead of 10) applied to one pathway of the adult hippocampal slices induced only STP. After incubation with BDNF for 3 h, the same weak tetanus applied to the second pathway resulted in a persistent LTP_T which averaged 34% as measured 40 min after the tetanic stimulation (right). Control slices incubated in ACSF for 4 h never showed LTP_T in response to the weak tetanus. **c**, Effects on LTP_K by BDNF in p12–13 slices and by TrkB-IgG in adult slices. In the example (left), TEA was applied to the slices for the period indicated by the bar. Percentages of potentiation above baseline values were measured 60 min after washout of TEA from recording chamber (right). No difference was found between control and drug-treated slices (*t*-test). **METHODS.** Extracellular recording and drug application are described in Fig. 1, except that a cut was made between CA1 and CA3 to prevent epileptiform activity^{19–21}. Schaffer–commissural afferents were stimulated at 0.05 Hz. After a stable baseline of EPSPs was established, TEA (25 mM) was perfused into the interface chamber for 10 min, while the presynaptic afferents were continuously stimulated at the same frequency. The application of TEA induced a sustained postsynaptic depolarization, bursts of Ca²⁺ spikes, and widespread increases in dendritic intracellular Ca²⁺ concentration while presynaptic activity was essentially unaffected²¹.

tion may explain why exogenous neurotrophins prevent the synaptic elimination that normally accompanies ocular dominance column formation²⁹, and rescue the effect of monocular deprivation³⁰ in the visual cortex. Our results may therefore provide insights into molecular mechanisms for the activity-dependent regulation of synaptogenesis and plasticity. □

Received 5 January; accepted 23 April 1996.

- Patterson, S., Grover, L. M., Schwartzkroin, P. A. & Bothwell, M. *Neuron* **9**, 1981–1088 (1992).
- Castren, E. *et al. NeuroReport* **4**, 895–898 (1993).
- Isackson, P. J. *Curr. Opin. Neurobiol.* **5**, 350–357 (1995).
- Maisonpierre, P. C. *et al. Neuron* **5**, 501–509 (1990).
- Friedman, W. J., Olson, L. & Persson, H. *Eur. J. Neurosci.* **3**, 688–697 (1991).
- Ringstedt, T., Kagercrantz, H. & Persson, H. *Dev. Brain Res.* **72**, 119–131 (1993).
- Dugich-Djordjevic, M. M., Ohsawa, F. & Heftli, F. *NeuroReport* **4**, 1091–1094 (1993).
- Harris, K. M. & Teyler, T. J. *J. Physiol., Lond.* **346**, 27–48 (1984).
- Muller, D., Oliver, M. & Lynch, G. *Dev. Brain Res.* **49**, 105–114 (1989).
- Jackson, P. S., Suppes, T. & Harris, K. M. *J. Neurophysiol.* **70**, 1412–1419 (1993).
- Shelton, D. L. *et al. J. Neurosci.* **15**, 477–491 (1995).
- Figurov, A., Boddeke, H. & Muller, D. *Eur. J. Neurosci.* **5**, 1035–1041 (1993).
- Kang, H. & Schuman, E. M. *Science* **267**, 1658–1662 (1995).
- Chao, M. V. *Neuron* **9**, 583–593 (1992).
- Ip, N. Y. *et al. Neuron* **10**, 137–149 (1993).
- Bliss, T. V. P. & Collingridge, G. L. *Nature* **361**, 31–39 (1993).
- Malenka, R. C. & Nicoll, R. A. *Trends Neurosci.* **16**, 521–527 (1993).
- Zucker, R. S. *A. Rev. Neurosci.* **12**, 13–31 (1989).
- Aniksztejn, L. & Ben-Ari, Y. *Nature* **349**, 67–69 (1991).
- Huang, Y. & Malenka, R. C. *J. Neurosci.* **13**, 568–576 (1993).
- Petrozzino, J. J. & Connor, J. A. *Hippocampus* **4**, 546–558 (1994).
- Lohoff, A. M., Ip, N. Y. & Poo, M. M. *Nature* **363**, 350–353 (1993).
- Knipper, M., Leung, L. S., Zhao, D. & Rylett, R. J. *NeuroReport* **5**, 2433–2436 (1994).
- Kim, H. G., Wang, T., Olafsson, P. & Lu, B. *Proc. natn. Acad. Sci. U.S.A.* **91**, 12341–12345 (1994).
- Wang, T., Xie, K. W. & Lu, B. *J. Neurosci.* **15**, 4796–4805 (1995).
- Levine, E. S., Dreyfus, C. F., Black, I. B. & Plummer, M. R. *Proc. natn. Acad. Sci. U.S.A.* **92**, 8074–8077 (1995).
- Thoenen, H. *Science* **270**, 593–596 (1995).
- Korte, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 8856–8860 (1995).
- Cabelli, R. J., Horn, A. & Shatz, C. J. *Science* **267**, 1662–1666 (1995).
- Riddle, D. R., Lo, D. C. & Katz, L. C. *Nature* **378**, 189–191 (1995).

ACKNOWLEDGEMENTS. We thank Genentech, Inc. (San Francisco) for providing neurotrophins; P. T. Kelly, K. M. Harris, J. A. Connor, M.-m. Poo, Y. Rao, M. Yuzaki, S. Udenfriend and S. Otani for discussions and for comments on the manuscript, and D. L. Shelton for the TrkB-IgG fusion protein. P.O. is supported by postdoctoral fellowships from the Roche Research Foundation and the Swiss National Science Foundation.

CORRESPONDENCE and requests for materials should be addressed to B.L. in Bethesda (e-mail: LUB@codon.nih.gov).

An RNA-dependent ATPase associated with U2/U6 snRNAs in pre-mRNA splicing

Deming Xu*, Shahrzad Nouraini*, Deborah Field*, Shou-Jiang Tang* & James D. Friesen*

Department of Genetics, Research Institute, Hospital for Sick Children, and Department of Molecular and Medical Genetics, University of Toronto, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

THE hydrolysis of ATP by a group of RNA-dependent ATPases (DEAD/H proteins¹) is required for spliceosome assembly, but not for the subsequent transesterification reactions². Little is known about the function of these ATPases in relation to the RNA conformational changes that occur in formation of active structures, in which U2/U6 small nuclear RNA (snRNA) interactions^{3,4} are essential for splicing to take place. Using a synthetic lethal genetic screen, we have isolated four yeast splicing factors involved in U2/U6 snRNA interactions (D.X. *et al.*, manuscript in preparation). The RNA-dependent ATPase activity associated with one such factor, the Slt22 protein, is stimulated preferentially by annealed U2/U6 snRNAs. Both mutant *slt22-1* and U2 snRNA cause a reduction in stimulation. The *slt22-1* mutation

* Present address: Banting and Best Department of Medical Research, University of Toronto, 112 College Street, Toronto, Ontario M5G 1L6, Canada.

blocks splicing at or before the first step, resulting in the accumulation of an unusual complex which lacks U5 snRNA. Our results indicate that the U2/U6 snRNA interactions facilitated by Slt22 are also involved in the interaction of U5 snRNA with the spliceosome.

Four chromosomal *slt* (for synthetic lethal with U2) mutations were isolated on the basis of synthetic lethality with a mutant U2 snRNA (*snr20-11nt*) (Fig. 1a, b) that could potentially perturb U2/U6 helix II interaction (D.X. *et al.*, manuscript in preparation)⁵. One of these, *slt22-1*, is also synthetic-lethal with *snr20-9nt* and *snr20-G21C* (ref. 14) (Fig. 1a), but is otherwise highly allele-specific with respect to other mutations in U2 (data not shown). The synthetic lethal phenotype arose from a single-gene mutation that also conferred temperature sensitivity ($\geq 33^\circ\text{C}$) and slow growth ($23\text{--}30^\circ\text{C}$) (Fig. 1b, and data not shown). The mutation itself caused a splicing defect in yeast extracts and blocked splicing at or before the first-step reaction (Fig. 1c).

The wild-type (WT) *SLT22* gene was cloned by complementation (Fig. 2a). It is identical to *YER172c* (Genbank accession number, P32639) and encodes a protein of 2,163 amino acids. Its central region (residues 500–900) contains motifs that are conserved in a class of RNA-dependent ATPase/RNA helicase-like proteins (data not shown). Slt22 is related distantly to DEAD (Prp5 and Prp28) and DEAH (Prp2, Prp16 and Prp22) ATPase/RNA helicase families^{1,2,5} in that it contains DEIH in ATPase box B and two additional motifs that are conserved in a class of relatively large ATPase/RNA helicase-like proteins (Fig. 2a and data not shown): P×KAL and GIG×HHA/GGL motifs. The C-terminal region (residues 1,349–2,163) shows homology (20% identity and 40% similarity) to the central ATPase region (residues 500–1,348), which also contains a putative leucine zipper (residues 1,073–1,115). The C-terminal region contains a second GKT (nucleotide-binding) box and a stretch of four amino acids, DDAH, which resembles ATPase box B (DEAH). Other ATPase/RNA helicase motifs are not conserved in this region. *SLT22* is an essential gene (data not shown). The mutation that corresponds to *slt22-1* (E909K) is located in the central ATPase/RNA helicase region and lies on the C-terminal side of the QM×GRAGR motif (Fig. 2a), shown to be important for RNA-dependent ATPase and RNA helicase activities for eIF4A (ref. 6).

A haemagglutinin (HA) epitope was introduced at the N termini of both *SLT22* and *slt22-1* genes carried on a 2- μm plasmid. The tagged proteins were expressed in an *SLT22* deletion yeast strain and were purified partially by immunoprecipitation. Similar amounts of tagged proteins were recovered from extracts of cells containing HA-Slt22 and HA-slt22-1 (but not from a wild-type extract lacking the HA-tagged proteins), as determined by silver staining and protein blotting of the immunoprecipitants (data not shown), indicating that the *in vivo* defect associated with *slt22-1* is not due to instability of the mutant protein.

HA-Slt22 immunoprecipitants (containing $\sim 10\text{ ng}$ Slt22) were assayed for ATPase activity stimulated by various snRNAs transcribed *in vitro*; U2, U2S (a short version of U2 containing stems I and II), U4, U5 and U6. Before the ATPase assay, these snRNAs (250 ng, in excess; Fig. 2c) were annealed separately or in mixtures (U2/U6, U2S/U6 and U4/U6). Three- to fourfold stimulation of activity was observed with

all the snRNAs tested. However, the ATPase activity was stimulated preferentially (~ 15 -fold) by U2/U6 and U2S/U6 snRNAs (Fig. 2b); this is about twofold greater than the sum of stimulation by U2 and U6 separately (Fig. 2c). A mutant U2 snRNA (U2(11ntS); Fig. 1a), used in the original genetic screening, was $\sim 25\%$ less active than wild-type U2S/U6 in stimulating ATPase activity (Fig. 2e). In contrast, U2S/U6-stimulated activity was not seen when immunoprecipitants containing similar amounts of

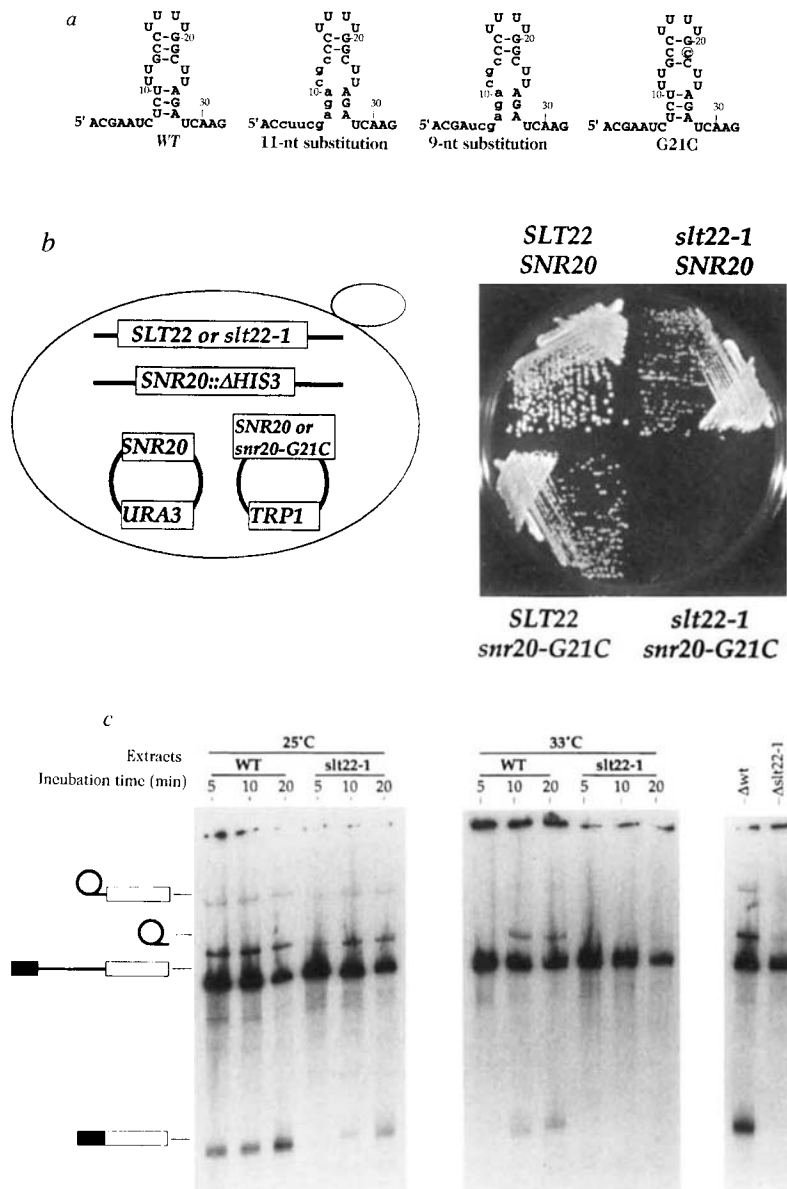


FIG. 1 Isolation and characterization of the *slt22-1* mutant strain. a, Structures of stem I of wild-type (WT) and three mutant U2 snRNAs (ref. 14) that are synthetically lethal with *slt22-1*. b, Synthetic lethality of *slt22-1* and *snr20-G21C*. Left, yeast strain in which synthetic lethality was tested; right, two-day growth (at 30°C) on medium containing 5-fluoroorotic acid of *SLT22* and *slt22-1* strains in combination with WT and mutant U2 snRNAs. c, *In vitro* splicing activity of WT and *slt22-1* extracts. The precursor, final products (mature RNA and lariat-intron) and intermediates (lariat-intron-exon2 and exon1) are indicated on the left side.

METHODS. For b, *S. cerevisiae* strains were derived from W303-1A or W303-1B (*Mata* or *Matx ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100*). The original *slt22-1* mutation was isolated on the basis of synthetic lethality using a strain described elsewhere¹⁴, following mutagenesis with ethylmethane sulphonate (in preparation). For c, whole-cell splicing extracts were prepared from WT and *slt22-1* cells according to ref. 15, with modifications. ^{32}P -labelled actin pre-mRNA was used as a substrate in a 20-min reaction with untreated extracts at 25°C (left) or 33°C (centre), or extracts pre-heated at 37°C (for 45 min, $\Delta\text{slt22-1}$ and ΔWT) before splicing at 25°C (right). Splicing products were resolved on a 5% denaturing polyacrylamide gel.

mutant *slt22-1* were used in the ATPase assay (Fig. 3d, e); ATPase activities stimulated by U2S, U2(11nt)S or U6 alone were almost comparable to those of wild-type immunoprecipitants; that is, three- to fourfold greater than the basal level (Fig. 3e). As no obvious changes in association with other proteins were observed in HA-*slt22-1* immunoprecipitants (data not shown), the observed U2S/U6 snRNA-stimulated ATPase activity is probably attributable to Slt22 itself. The rate of ATP hydrolysis under stimulation by U2S/U6 was estimated to be ~ 45 mol P_i per min per mol Slt22. The reduction in the U2S/U6-stimulated ATPase activity of HA-

slt22-1 can be attributed directly to the E909K mutation in the central ATPase region of the protein.

We examined spliceosome assembly in a *slt22-1* extract incubated at the non-permissive temperature. The assembly of splicing complexes at 25 and 33 °C in a wild-type extract and in an *slt22-1* extract at 25 °C (Fig. 3a, lanes 1–12) was consistent with previous observations⁷, but an unusual complex was detected at 33 °C in the *slt22-1* extract (complex X; Fig. 3a, lanes 13–16). This complex appeared later than complexes B and A2 (probably A2–1; ref. 7), as it was not present at very early times (Fig. 3a, lane 13). After

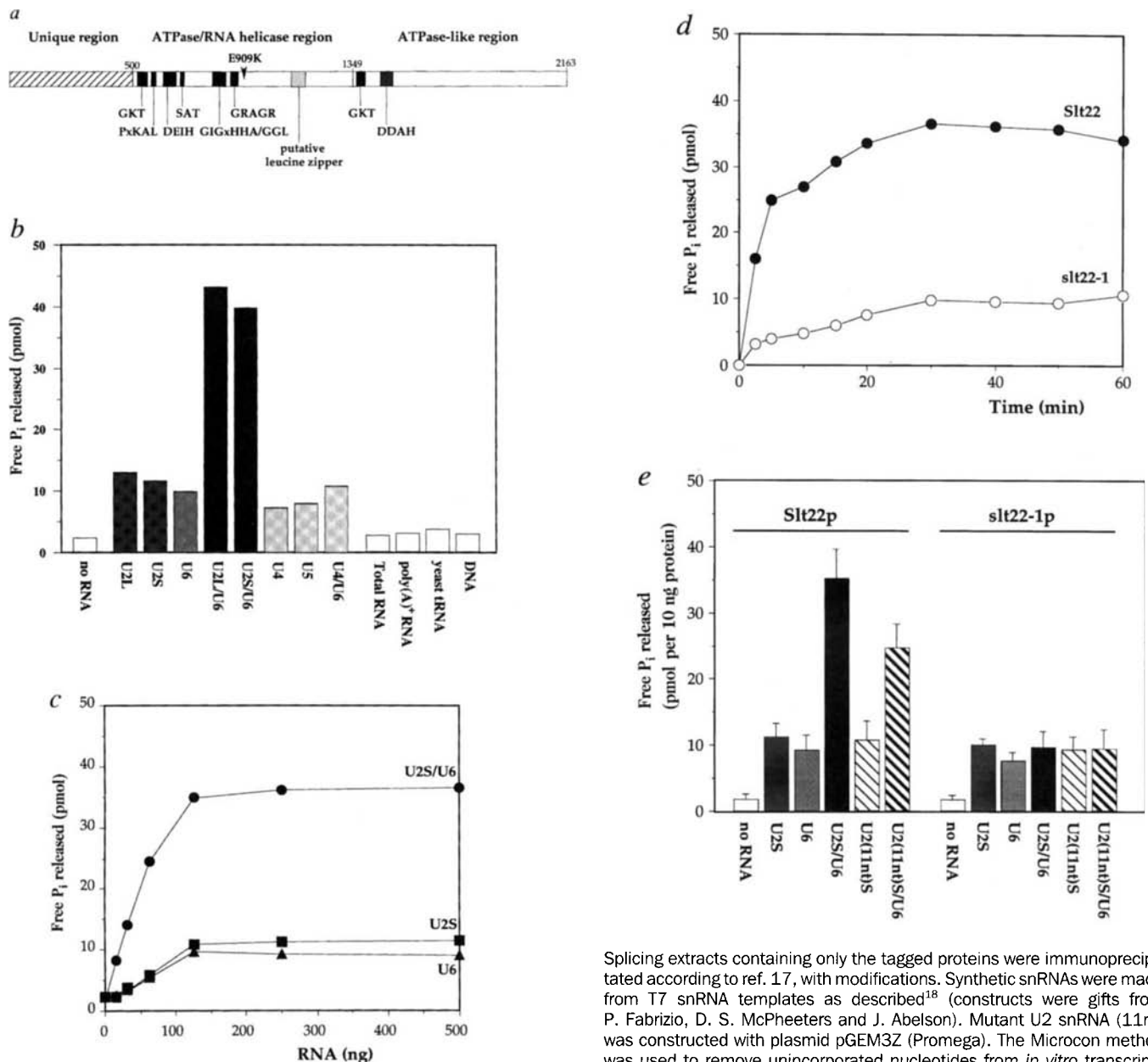


FIG. 2 Molecular cloning of Slt22 and RNA-dependent ATPase assays. **a**, Schematic of Slt22: filled boxes indicate conserved motifs described in text. **b**, Stimulation by different RNAs or by T7-U2 plasmid DNA of ATPase activities associated with HA-Slt22. **c**, U2S/U6 snRNA-stimulated ATPase activities as a function of RNA concentration. **d**, Time course of U2S/U6 snRNA-stimulated ATPase activities associated with HA-Slt22 and HA-Slt22-1. **e**, Stimulation by WT and mutant U2S/U6 snRNAs of ATPase activities associated with HA-Slt22 and HA-Slt22-1. **METHODS.** **a**, Slt22 was cloned by complementation¹⁶ of the temperature-sensitive (ts) phenotype of *slt22-1*; **b**, **c** and **d**, an HA epitope was introduced at the fifth amino acid of the protein. The modified Slt22 and *slt22-1* genes were carried on a pRS424 plasmid in an Slt22 Δ strain.

Splicing extracts containing only the tagged proteins were immunoprecipitated according to ref. 17, with modifications. Synthetic snRNAs were made from T7 snRNA templates as described¹⁸ (constructs were gifts from P. Fabrizio, D. S. McPheeters and J. Abelson). Mutant U2 snRNA (11nt) was constructed with plasmid pGEM3Z (Promega). The Microcon method was used to remove unincorporated nucleotides from *in vitro* transcripts (Amicon). Immunoprecipitants containing ~ 10 ng tagged proteins were used in a single ATPase assay¹⁹ with 50 mM Tris-HCl (pH 7.5), 5 mM MgCl₂, 1 mM DTT, 2 μ M ATP and 12.5 nCi [γ -³²P]ATP in a final volume of 25 μ l containing 250 ng RNA as indicated (**b**, **d** and **e**) or various amount of U2S, U6 and U2S/U6 snRNAs (**c**). Reaction mixtures were incubated at 25 °C for 20 min, or as indicated in **d**, with shaking. Aliquots (1.5 μ l) were loaded on polyethyleneimine cellulose plates and developed as described¹⁹. ATP hydrolysis, measured by release of free phosphate, was calculated from the percentage of free radioactive phosphate released, and background free phosphate resulting from incubation with immunoprecipitants of untagged WT extract was subtracted. **e**, Average ATPase activities obtained from at least three independent experiments; bars indicate standard deviations.

incubation at 33 °C for 20 min (Fig. 3b), three complexes were present in the mutant extract. The complex with lowest mobility contained U2, U5 and U6 snRNAs, but not U4 snRNA (Fig. 3c), and thus is related to complex A1. However, only U2 and U6 snRNAs were detected in complex X (Fig. 3c). The complex labelled Y showed very similar mobility to that of the H complexes (Fig. 3a) noted previously⁷ and has not been investigated further. These complexes (A1-like, X and Y), once formed, were unable to return to the normal splicing pathway and regain catalytic activity (Fig. 3d).

Slt22 may serve a proof-reading function for U2/U6 snRNA interactions. Its action may be similar to that proposed for Prp16 (refs 8,9), in which the rate of ATP hydrolysis, stimulated by RNA substrate(s) with the correct conformation, determines the choice between productive and default pathway. All three mutant U2

snRNAs used in this study (Fig. 1a) have the potential to form alternative base-pair interactions with wild-type U6 snRNA (Fig. 4b), which may perturb structures important for splicing reaction. One of them, U2(11nt), results in reduced stimulation of ATPase activity. The synthetic lethality may thus be due to exacerbation of impaired interactions between a mutant RNA-dependent ATPase (slt22-1) and a target {U2(11nt)/U6 snRNA} bearing mutations that can form an abnormal structure. Complex X, lacking U5 snRNA, may be the product of the default pathway, triggered by a malfunctioning slt22-1, and derived from other normal complexes, probably at the transition from A1 to A2-2; that is, the formation of active spliceosomes. Alternatively, this complex may represent a trapped form of a transient state in the normal splicing pathway, which is not observable otherwise. In either case, the formation of complex X in the *slt22-1* extract

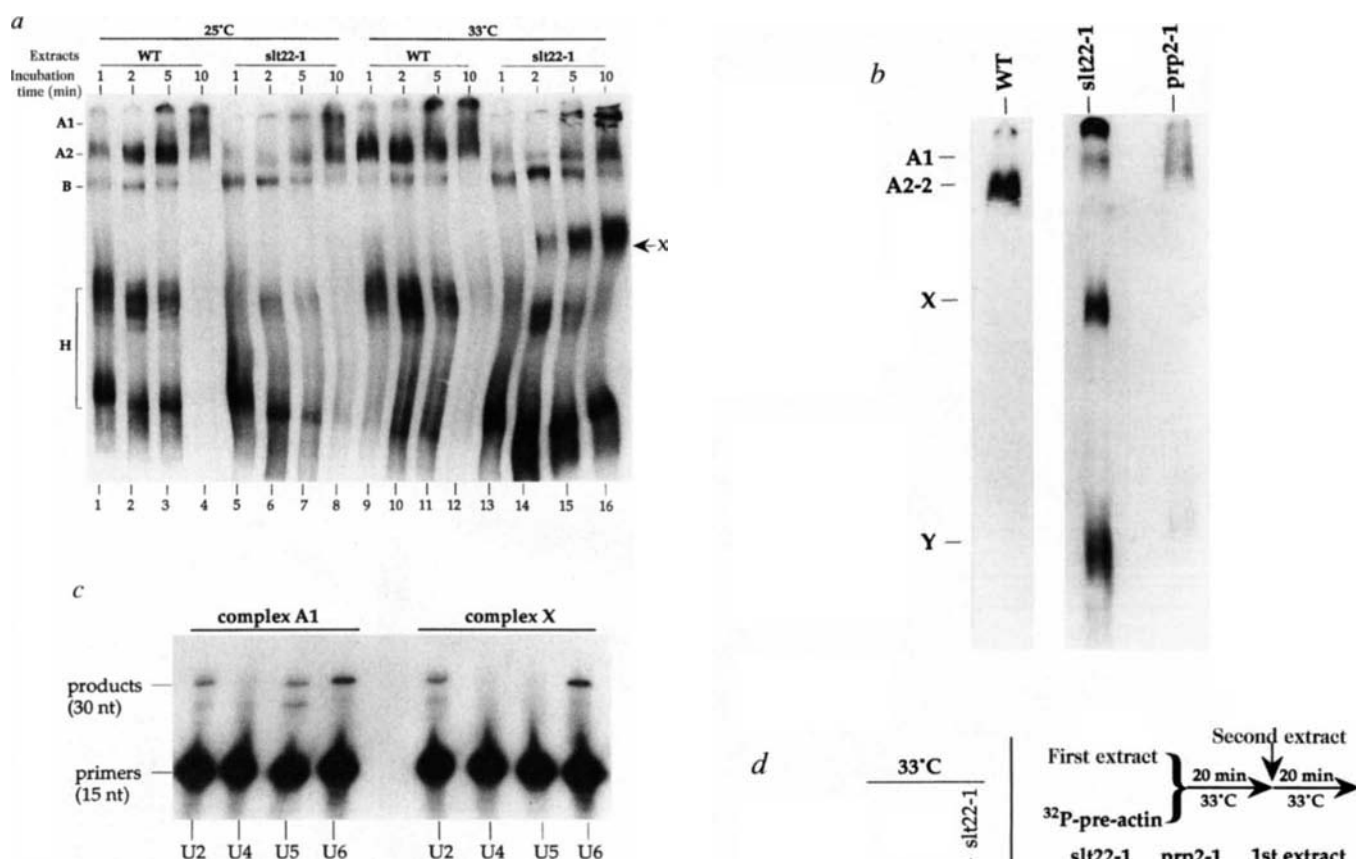
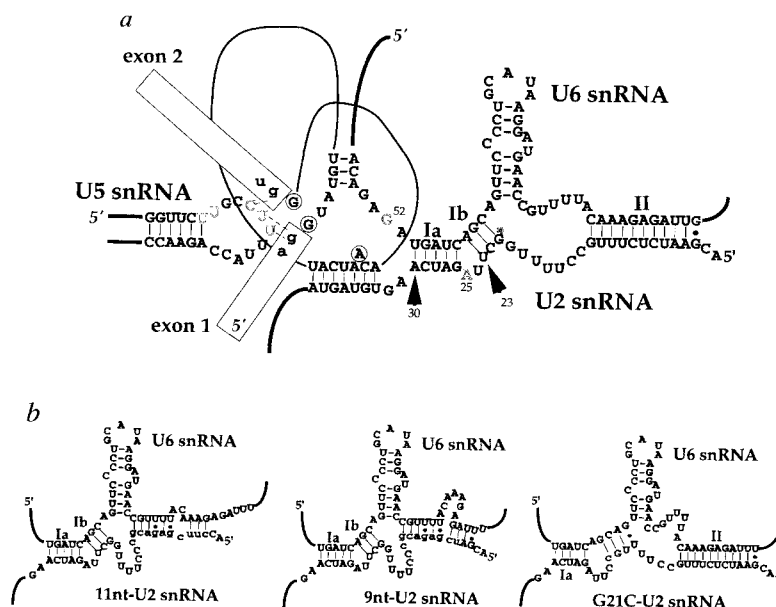


FIG. 3 Formation of a new splicing complex in a *slt22-1* extract. **a**, Native-gel kinetic analysis of spliceosome assembly in WT and *slt22-1* extracts incubated at 25 and 33 °C. The designation A2 includes complexes A2-1 and A2-2 (ref. 7), which are indistinguishable in this assay. **b**, Native-gel analysis of spliceosome assembly in WT, *slt22-1* and *prp2-1* extracts following 20 min incubation at 33 °C, showing the final complexes. **c**, snRNA content of complexes A1-like and X. The snRNA content of the normal spliceosome complexes has been described^{7,20}. **d**, Rescue of splicing activity. Left, standard splicing reactions were carried out at 33 °C with WT, *prp2-1*, *slt22-1* extracts, or *prp2-1* plus *slt22-1* extracts mixed before addition of substrate (lanes 1–4); right, to test whether final complexes formed in the *slt22-1* extract at 33 °C can be reintroduced into normal pathway to gain splicing activity, splicing reactions with *slt22-1* (lanes 5 and 6) and *prp2-1* (as a control, lanes 7 and 8) extracts were allowed to proceed at 33 °C for 20 min, then aliquots of second extracts (1/2 volume) were added. Reactions were kept at 33 °C for a further 20 min. METHODS. **a**, **b** and **c**, Splicing reactions are described in Fig. 1 legend. Native-gel electrophoresis was done according to ref. 7, with modifications. The amount of substrate used in **b** and **d** was one-quarter of that used in **a** and in Fig. 1 in order to ensure that all of it was incorporated into complexes during the initial part of the experiment. **c**, RNA was extracted from strips of the native gel containing complexes A1-like and X that had been formed

using non-radioactive pre-mRNA substrate at 33 °C. This RNA was subjected to primer-extension analysis using primers specific to nucleotides 16–30 of snRNAs indicated.

FIG. 4 U2/U6 snRNA interactions and the formation of possible alternative structures. *a*, U2/U6 snRNA interactions. Intermolecular U2/U6 snRNA interactions, in which conserved nucleotides involved in either or both splicing reactions^{21–23} are brought into close proximity, are important in splicing. Two alternative U2/U6 snRNA structures have been proposed^{3,4}; both include helices I and II. The structure proposed by Madhani and Guthrie³ is shown. Nucleotides involved directly in the transesterification reaction are circled. Highlighted residues in the invariant loop I of U5 snRNA²⁴ are crosslinked to residues in both exons²⁵. U23 and A30 (arrowheads) in U2 snRNA are also crosslinked to exon 2 (ref. 25). Another known tertiary interaction includes that between G52 in U6 snRNA and A25 in U2 snRNA²⁶. Asterisk indicates the G21 residue in U2 snRNA. *b*, Formation of possible alternative U2/U6 snRNA structures by mutations in U2 snRNA. The three U2 snRNA mutants that are synthetically lethal with *slt22-1* may form abnormal U2/U6 structures in which helix II (in the cases of 11-nt and 9-nt substitutions) or helix Ib (in the case of G21C) is distorted. These U2 snRNA mutants are also synthetically lethal with factors that are important for U5 snRNA function (see text).



suggests hitherto-unsuspected interactions between U2 and/or U2/U6 snRNA and U5 snRNA in the mature spliceosome.

Sl22 has recently been found to be associated with the U4/U6.U5 tri-snRNP (ref. 10). Our genetic screen (manuscript in preparation) has revealed that both the 11- and 9-nucleotide substitutions of U2 snRNA, which affect the viability of yeast¹⁴, are synthetically lethal with mutations in at least four protein factors, namely Sl22, Sl11, Sl21/Prp8 and Sl17/Slu7 (our unpublished observations). In addition, mutations in U2 snRNA are also synthetically lethal with *slu4-1/prp17* (refs 11,12) and *slu5-1* (ref. 11, and our unpublished observations). Almost all these factors are known to be related functionally to U5 snRNA^{11,13}, which supports our observation that spliceosomal U5 snRNA depends on a functional Sl22. Thus the integrity of the U2/U6 snRNA interactions (Fig. 4), monitored by Sl22 and in association with these other factors, is important for U5 snRNA function in both steps of the transesterification reactions. □

Uncoupling of S phase and mitosis induced by anticancer agents in cells lacking p21

Todd Waldman, Christoph Lengauer, Kenneth W. Kinzler & Bert Vogelstein

The Howard Hughes Medical Institute, Johns Hopkins Oncology Center, and Program in Human Genetics, 424 N. Bond Street, Baltimore, Maryland 21231, USA

PRECISE coordination of the S and M phases of the eukaryotic cell cycle is critical not only for normal cell division, but also for effective growth arrest under conditions of stress. When damaged, a cell must communicate signals to both the mitotic and DNA synthesis machineries so that a mitotic block is not followed by an extra S phase, or vice versa. The biochemical mechanisms regulating this coordination, termed checkpoints, have been identified in lower eukaryotes, but are largely unknown in mammalian cells^{1–3}. Here we show that p21^{WAF1/CIP1}, the prototype inhibitor of cyclin-dependent kinases (CDKs)⁴, is required for this coordination in human cells. In the absence of p21, DNA-damaged cells arrest in a G2-like state, but then undergo additional S phases without intervening normal mitoses. They thereby acquire grossly deformed, polyploid nuclei and subsequently die through apoptosis. Perhaps not by coincidence, the DNA-damaging agents that can cause S/M uncoupling are used in the clinic to kill cancer cells preferentially.

Our experiments were initially done with isogenic, diploid lines derived from the human colorectal cancer cell line HCT116 following homologous recombination to delete the endogenous p21 genes⁵. When treated with adriamycin, the parental (p21^{+/+}) cells remained attached to the plate and appeared morphologically normal, as do other colorectal epithelial cell lines that growth arrest following DNA damage³. In contrast, p21-deficient cells (p21^{-/-}) shrank, rounded, and detached from the dish, suggesting apoptosis had occurred (Fig. 1a, b). To confirm this suggestion, cells were stained with the DNA-binding dye H33258, revealing condensed chromatin and fragmented nuclear morphology in the

Received 21 December 1995; accepted 25 April 1996.

- Schmid, S. R. & Linder, P. *Molec. Microbiol.* **6**, 283–292 (1992).
- Moore, M. J. et al. in *The RNA World* (eds Gesteland, R. F. & Atkins, J. F.) 303–357 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1993).
- Madhani, H. D. & Guthrie, C. *Cell* **71**, 803–817 (1992).
- Sun, J.-S. & Manley, J. L. *Genes Dev.* **9**, 843–854 (1995).
- Madhani, H. D. & Guthrie, C. *A. Rev. Genet.* **28**, 1–26 (1994).
- Pause, A., Methot, N. & Sonenberg, N. *Molec. cell. Biol.* **13**, 6789–6798 (1993).
- Cheng, S. C. & Abelson, J. *Genes Dev.* **1**, 1014–1027 (1987).
- Burgess, S. M. & Guthrie, C. *Cell* **73**, 1377–1391 (1993).
- Burgess, S. M. & Guthrie, C. *Trends biochem. Sci.* **18**, 381–384 (1993).
- Lauber, J. et al. *EMBO J.* (in the press).
- Frank, D., Patterson, B. & Guthrie, C. *Molec. cell. Biol.* **12**, 5197–5205 (1992).
- Jones, M. H., Frank, D. N. & Guthrie, C. *Proc. natn. Acad. Sci. U.S.A.* **92**, 9687–9691 (1995).
- Brown, J. D. & Beggs, J. D. *EMBO J.* **11**, 3721–3729 (1992).
- Field, D. & Friesen, J. D. *Genes Dev.* **10**, 489–501 (1996).
- Lin, R.-J. et al. *J. Biol. Chem.* **260**, 14780–14792 (1985).
- Rose, M. D. & Broach, J. R. *Meth. Enzym.* **194**, 195–230 (1991).
- Kolodzie, P. A. & Young, R. *Meth. Enzym.* **194**, 508–519 (1991).
- McPheeters, D. S. & Abelson, J. *Cell* **71**, 819–831 (1992).
- Fuller-Pace, F. V. et al. *EMBO J.* **12**, 3619–3626 (1993).
- Tam, W.-Y. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 10821–10825 (1993).
- Fabrizio, P. & Abelson, J. *Science* **250**, 404–409 (1990).
- Madhani, H. D. et al. *Genes Dev.* **4**, 2264–2277 (1990).
- McPhoeters, D. S. & Abelson, J. et al. *Cell* **71**, 819–831 (1992).
- Frank, D. N., Roeha, H. & Guthrie, C. *Molec. cell. Biol.* **14**, 2180–2190 (1994).
- Newman, A. J., Teigelkamp, S. & Beggs, J. *D. RNA* **1**, 968–980 (1995).
- Madhani, H. D. & Guthrie, C. *Genes Dev.* **8**, 1071–1086 (1994).

ACKNOWLEDGEMENTS. We thank F. Dietrich, R. Davis, S. Noble, Y. Wang, C. Guthrie, D. Fushman, M. Ohno, S. Teigelkamp and R. Lührmann for sharing unpublished results, V. Stambolic and J. Woodgett for the 12CA5 cell line; R. Slaaby and A. Moris for technical assistance; and B. Andrews, T. Yager, D. Jansma, I. Donaldson and V. Lay for their comments. This research was supported by the Medical Research Council of Canada and the National Cancer Institute of Canada.

CORRESPONDENCE and requests for materials should be addressed to J.D.F. (e-mail: jim@resunix.ri.sickkids.on.ca).

p21^{-/-} cells, whereas the p21^{+/+} cell nuclei were homogeneous and round (Fig. 1c, d). The condensed, fragmented nuclei were stained intensely by TUNEL⁶⁻⁷, which detects the presence of DNA breaks characteristic of apoptosis (Fig. 1e, f). Time-course analysis demonstrated that apoptosis was complete between 60 and 90 hours following adriamycin treatment (Fig. 2a). The response to adriamycin was found to be identical in a second, independently isolated p21^{-/-} clone (Fig. 2a).

To determine whether these observations were generalizable with respect to DNA-damaging agents, we treated the cells with γ -irradiation, the topoisomerase-I inhibitor camptothecin, and the topoisomerase-II inhibitor etoposide. After etoposide treatment or irradiation, the parental p21^{+/+} cells remained healthy whereas their p21^{-/-} derivatives died, as judged by phase-contrast microscopy and H33258 staining (Fig. 2b, c). Adriamycin and etoposide induced apoptosis in p21-deficient cells slightly faster (complete by 90 h) than irradiation (complete by 120 h). With camptothecin, the parental p21^{+/+} cells eventually underwent apoptosis, but this was significantly delayed compared to p21-deficient cells (Fig. 2d). Cells with one normal copy of the p21 gene and one deleted copy (p21^{+/-}) behaved similarly to the parental cells after treatment with adriamycin and etoposide (Fig. 2a, b), whereas treatment with camptothecin or γ -irradiation revealed a heterozygote effect (Fig. 2c, d).

Flow cytometry demonstrated that specific cell-cycle changes were associated with this apoptotic process. Following 30 h of exposure to adriamycin, p21^{+/+} cells were blocked in G1 and G2 phases, with few cells in S phase (Fig. 3c). In contrast, no G1 block was evident in the p21^{-/-} cells⁵, so that a nearly pure population of G2-arrested cells was observed (Fig. 3d). With longer treatments, the flow cytometry profile of p21^{+/+} cells remained largely unchanged (Fig. 3e, g), indicating a stable growth arrest⁸, whereas p21^{-/-} cells began to accumulate DNA in excess of 4C, then died (Fig. 3f, h). This characteristic phenotype—a block in G2, followed by additional rounds of DNA synthesis in the absence of mitosis—was observed in p21^{-/-} cells following treatment with each of the DNA-damaging drugs (for examples, see Fig. 3f, h, n, p). Following γ -irradiation, there was a particularly striking accumulation of polyploid cells, with 33% and 15% of nuclei exhibiting DNA contents of 8C and 16C, respectively (Fig. 3l).

Fluorescence *in situ* hybridization (FISH) allowed us to extend the flow cytometry results by linking abnormal nuclear morphology with DNA synthesis and polyploidy in individual cells. Nuclei became lobulated 10–30 h before apoptosis in p21-deficient cells treated with DNA-damaging drugs or radiation. For example, 31% of p21^{-/-} cell nuclei were lobulated following 32 h of adriamycin treatment (Fig. 4a). Transmission electron microscopy

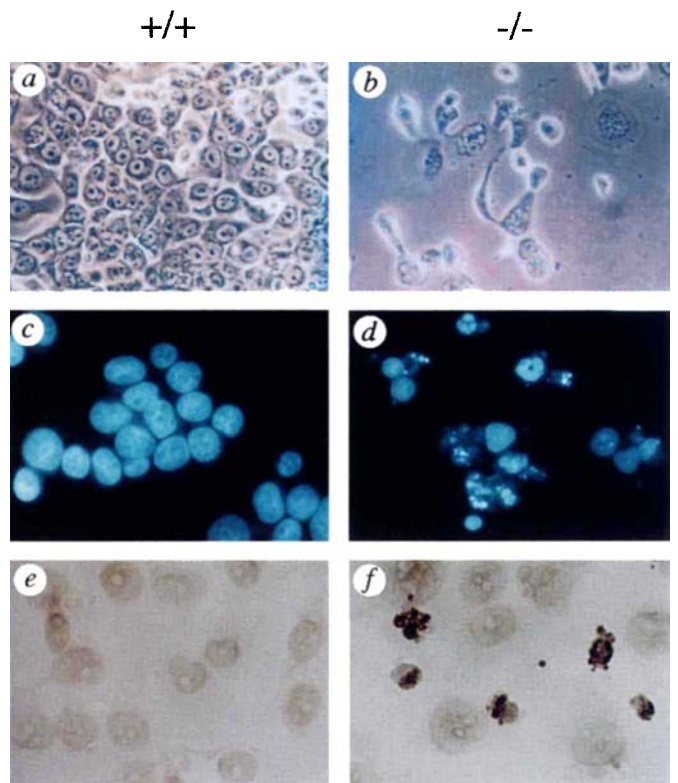
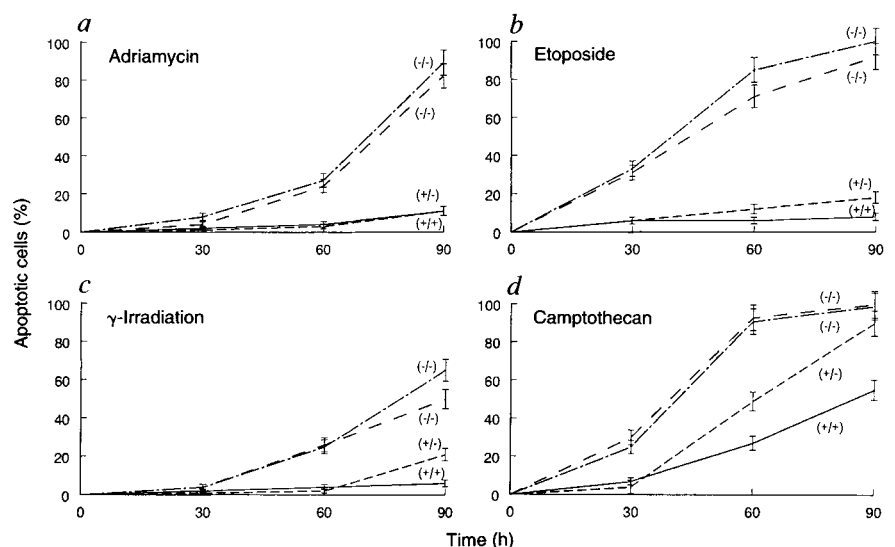


FIG. 1 Apoptosis in p21-deficient cells. Cells were grown for 90 h in the presence of adriamycin and viewed with phase-contrast microscopy (a, b), after staining with the DNA-binding dye H33258 (c, d), or after staining with the TUNEL assay to detect fragmented DNA (e, f).

METHODS. HCT116 cells with (+/+) or without (-/-) p21 genes were generated by homologous recombination⁵. Logarithmically growing cells at 70% confluence in McCoy's 5A medium with 10% fetal calf serum were treated with adriamycin (0.2 $\mu\text{g ml}^{-1}$) for 90 h. For c and d, adherent cells were collected by trypsinization and combined with cells floating in the medium. After washing with Hanks buffered saline (HBS; Life Technologies), cells were resuspended in 40 μl HBS and added to 360 μl of a solution containing 0.7% NP-40, 4.7% formaldehyde and 11 $\mu\text{g ml}^{-1}$ H33258 in phosphate-buffered saline (PBS). Cells were then viewed under UV excitation and photographed with a Nikon Labophot microscope. For e and f, adherent cells were fixed in 1% formaldehyde, then incubated in the presence of exogenous terminal transferase and biotin-11-dUTP as described⁷. Labelled cells were detected by immunoperoxidase staining (Vector).

FIG. 2 Kinetics of apoptosis following drug or radiation treatment. Cells were continuously treated with a, adriamycin (0.2 $\mu\text{g ml}^{-1}$); b, etoposide (5 $\mu\text{g ml}^{-1}$); c, camptothecin (0.1 $\mu\text{g ml}^{-1}$); or d, irradiated (12 Gy). At the indicated times, cells were collected, stained with the DNA-binding dye H33258, and viewed by fluorescence microscopy to determine the fraction of apoptotic cells.

METHODS. Cells were grown, fixed, and stained as described in the legend to Fig. 1. Irradiation was delivered at 1 Gy min^{-1} using a ¹³⁷Cs source. Apoptotic cells were recognized as condensed, punctate nuclei (examples are shown in Fig. 1d) or ghosts with faintly stained, degrading nuclei. At least 200 cells were counted for each time point; the experiment was repeated with results virtually identical to those shown. Error bars represent the Poisson standard deviation determined by taking the square-root of counted events and converting it to per cent abundance. All counting was done in a blinded fashion. The (+/-) cells contained one normal p21 allele and one allele deleted by homologous recombination⁵.



provided high-resolution imaging of the nuclear lobulations (Fig. 4e) and revealed that at least some were physically connected by a single continuous sheath of nuclear membrane (not shown). Moreover, 96% of these lobulated nuclei were actively synthesizing DNA, as assessed by bromodeoxyuridine (BrdU) incorporation (Fig. 4b). FISH with a chromosome-3p probe showed that 4% of the lobulated nuclei contained 32 copies of chromosome 3, 4% contained 16 copies, 76% contained 8 copies, and 16% contained 4 copies (an example is shown in Fig. 4a). In contrast, none of the $p21^{+/+}$ cell nuclei was lobulated, only 5% incorporated BrdU, and all contained either 2 or 4 copies of chromosome 3, as expected for diploid cells such as HCT116 blocked in G1 or G2 (for examples, see Fig. 4c, d). FISH analysis with chromosome 7, 12, 17, and 18 probes confirmed the 8–16C DNA content in the $p21^{-/-}$ cells and normal DNA content in the parental cells, consistent with the flow-cytometry results shown in Fig. 3.

We next investigated whether the striking S/M uncoupling observed in the $p21^{-/-}$ cells was evident in tumour cells that had not been genetically engineered *in vitro*. Though mutations in $p21$ occasionally occur in tumours *in vivo*, no human cancer cell lines with such mutations have been reported⁹. We therefore examined DLD-1 cells, a colorectal cancer line with a nearly diploid DNA content like that of HCT116. Unlike HCT116 cells, however, DLD-1 cells contain mutated $p53$ genes¹⁰. The $p53$ protein has been shown to bind avidly to two sites within the $p21$ promoter¹¹, and $p53$ is necessary for the transcriptional activation of $p21$ following DNA damage¹². We therefore reasoned that $p53$ -mutant cells would not express $p21$ following DNA damage, and would be functionally equivalent to $p21^{-/-}$ cells. After treatment with radiation or adriamycin, DLD1 cells expressed little $p21$ (not shown), and underwent phenotypic changes similar to those observed in HCT116 $p21^{-/-}$ cells. For example, 48 h after irradiation, 26% of the nuclei were lobulated, and the majority (81%) of these lobulated nuclei were actively synthesizing DNA, as assessed by BrdU incorporation. FISH analysis showed that 32% of the treated cells contained 8 or more copies of chromosome 12 (an example is shown in Fig. 4f). Of the cells with lobulated nuclei, 24% contained 16 copies of chromosome 12, 62% contained 8 copies, and only 14% contained 4 copies. These nuclear changes were followed by apoptotic death. In contrast, none of the untreated, exponentially growing DLD1 cell nuclei was lobulated, and none contained more than four copies of chromosome 12. Finally, we tested three other colorectal cancer cell lines with

mutant $p53$ genes. Although these lines were aneuploid, making analysis more difficult, it was clear that S/M uncoupling following irradiation occurred in at least two of them (Fig. 4g–j).

These data provide unambiguous evidence supporting the hypothesis that disruption of a cell cycle checkpoint can have dramatic effects on the response to anticancer agents in mammalian cells^{1–3,13,14}. Although uncoupling of cell-cycle phases has been noted previously, the S/M uncoupling observed here has no precise correlate in lower eukaryotes^{15,16}. Our results clearly show that, following DNA damage, human cells lacking $p21$ or its upstream regulator $p53$ can initiate and often complete entire rounds of S phase in the absence of mitosis, leading to gross nuclear abnormalities and culminating in apoptosis. The checkpoint(s) keeping cells with damaged DNA from entering mitosis are functional in these cells, whereas the checkpoint that would normally prevent them from reinitiating S phase is defective. Furthermore, although the results obtained with the $p53$ -mutant cells cannot be used to identify the macromolecules regulating this checkpoint, the experiments with HCT116 cell derivatives unambiguously point to $p21$ as its required arbiter. In view of the known biochemical properties of $p21$, the enzymatic mediators of this checkpoint are undoubtedly cyclin-dependent kinases^{4,17,18}.

At least two more general implications can be drawn from these results. First, the S/M uncoupling observed here in response to DNA-damaging agents may also occur, to lesser extents, under physiological conditions *in vivo*. In has been shown, for example, that hypoxia within poorly vascularized portions of tumours results in the induction of $p53$ (ref. 19). Cells with $p53$ mutations may not only escape the growth inhibition normally conferred by $p53$, but may coincidentally endoreduplicate their genomes as a result of S/M uncoupling. Such endoreduplication may represent a first step in aneuploidy, giving rise to the chromosomal variability that is found in virtually all solid tumours^{20–23}. Second, the results are relevant to chemotherapeutic sensitivities. Such relative sensitivities are complex, involving differences in drug uptake and metabolism, the capacity to endure a prolonged state of growth arrest, and the susceptibility to apoptosis²⁴. Consequently, it is not surprising that the responses to such drugs vary among cell types and species. Nevertheless, our results demonstrate that, under defined conditions *in vitro*, S/M uncoupling can be one of the factors that can substantially affect this sensitivity. Thus, the experimental system described here suggests a potentially useful means of screening for new chemotherapeutic agents. Novel drugs

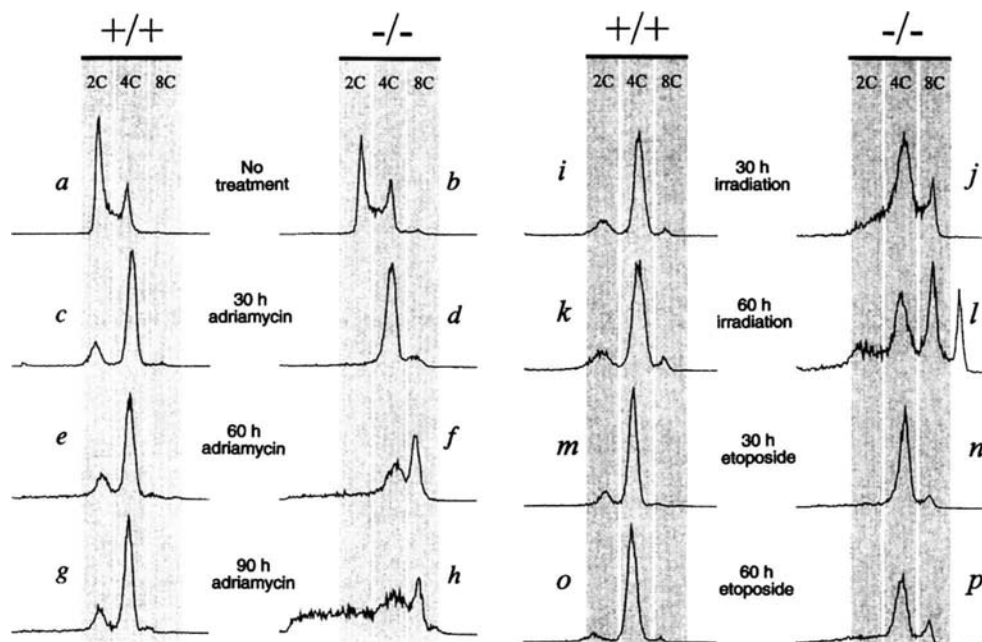


FIG. 3 Cell-cycle analysis of treated cells. Cells with (+/+) or without (-/-) intact $p21$ genes were stained following various periods of treatment and examined by flow cytometry. The cells were untreated (a, b), treated with adriamycin (c–h) or etoposide (m–p) for the indicated times, or treated with γ -irradiation (i–l) and examined 30–60 h later.

METHODS. Cell growth, drug or radiation treatment, fixation, and staining with H33258 are described in Fig. 2 legend. Flow cytometry was performed as before⁵. The G1 DNA content of the cells is denoted as 2C, and DNA content is plotted on a logarithmic abscissa.

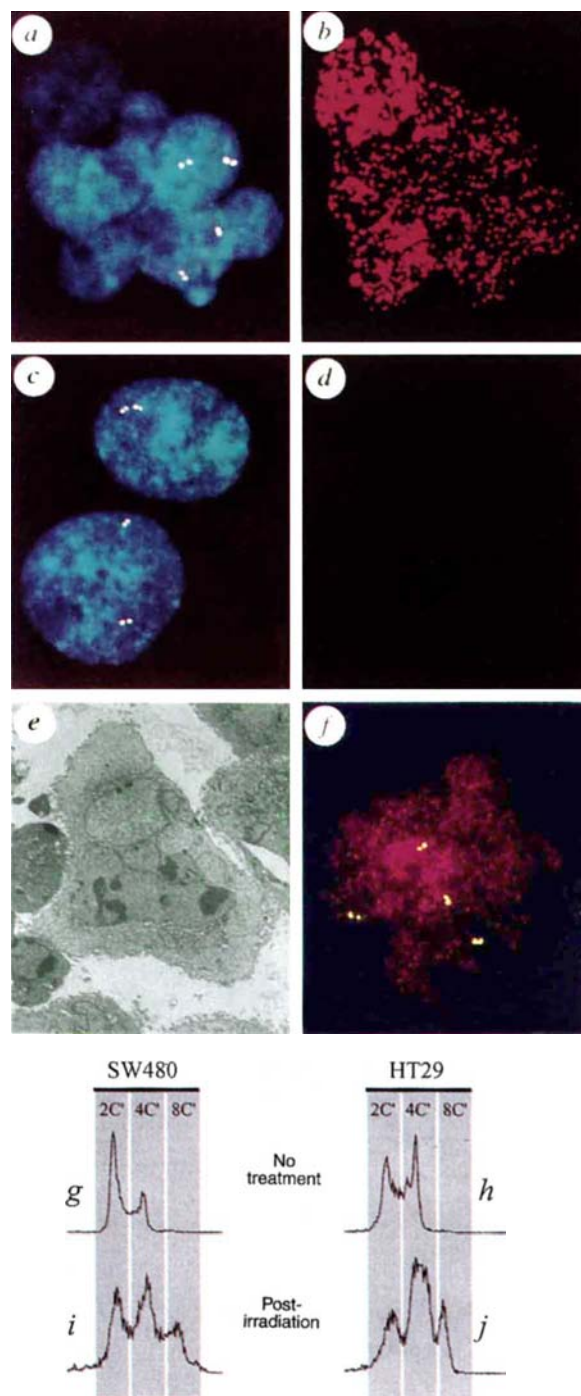


FIG. 4 Nuclear changes during S/m uncoupling. HCT116 cells without (a, b) or with (c, d) intact *p21* genes were treated with adriamycin, pulse-labelled with BrdU to assess DNA synthesis, fixed and hybridized with a chromosome-3 probe by FISH (a, c), then stained with the DNA-binding dye DAPI (a, c) and an anti-BrdU antibody (b, d). FISH signals are visualized as white dots; nuclear morphology is revealed by the blue DAPI stain (a, c). BrdU incorporation of the same cells is visualized in red (b, d). Note the single, lobulated nucleus in a and b. After irradiation, HCT116 *p21*^{-/-} cells were fixed, mounted and stained, then viewed by transmission electron microscopy at 2,000 $\times$ magnification (e). DLD1 cells were pulse-labelled with BrdU 48 h after irradiation, then fixed and analysed by FISH using a chromosome-12 probe, and stained with an anti-BrdU antibody (f). FISH signals are seen as yellow dots; BrdU incorporation is revealed in red. HT29 and SW480 cells were examined by flow cytometry, either in the untreated state (g, h) or after γ -irradiation (i, j).

METHODS. Cells were grown and treated with adriamycin as for Fig. 1. After 60 h incubation, cells were pulse-labelled with BrdU (5 μ M) for 1.5 h. The cells were then collected, fixed on slides with methanol:acetic acid (3:1), treated with RNase and pepsin²⁵, and FISH performed²⁶ with a P1 clone derived from 3p21.1-3. The P1-probe DNA was biotinylated by nick-translation and detected with fluorescein isothiocyanate (FITC) conjugated to avidin²⁷. BrdU incorporation was detected by indirect immunofluorescence using an anti-BrdU antibody (5 μ g ml⁻¹; Pharmingen) and an anti-mouse IgG TRITC conjugate (Sigma). Cells were counterstained with 0.05 μ g ml⁻¹ DAPI. Photographs were taken using a CCD camera (Photometrics) mounted on a Zeiss AxioPhot microscope. Digital image acquisition and processing was done as described²⁶. Cells were grown and treated with 12 Gy irradiation as for Fig. 2. HCT116 *p21*^{-/-} cells were collected 60 h after irradiation, microwave-pulse-fixed in 2% paraformaldehyde, 1% glutaraldehyde, 50 mM phosphate buffer, pH 7.3, and embedded in Spurr's Resin. 90–100-nm sections were then cut on a diatome, double-stained with uranyl acetate and lead citrate, and viewed on a Zeiss 10B transmission electron microscope operating at 60 kV (ref. 28). DLD1 cells were collected 48 h after irradiation and FISH was done with a P1 clone derived from 12p12–13.1. HT29 and SW480 cells were collected 34 or 60 h after irradiation, respectively, then fixed, stained with H33258, and analysed by flow cytometry as for Fig. 3. G1 DNA content of aneuploid cell populations is denoted as 2C.

that stimulate S/M uncoupling in *p21*^{-/-} but not in *p21*^{+/-} cells, as assessed by DNA synthesis or apoptosis, might prove as useful in the clinic as those tested here. ☐

Received 1st April; accepted 19 April 1996.

1. Murray, A. W. *Nature* **359**, 599–604 (1992).
2. Nurse, P. *Cell* **79**, 547–550 (1994).
3. Hartwell, L. H. & Kastan, M. B. *Science* **266**, 1821–1828 (1994).
4. Nasmyth, K. & Hunt, T. *Nature* **366**, 634–635 (1993).
5. Waldman, T., Kinzler, K. W. & Vogelstein, B. *Cancer Res.* **55**, 5187–5190 (1995).
6. Gavriel, Y., Sherman, Y. & Ben-Sasson, S. A. *J. Cell Biol.* **119**, 493–501 (1992).
7. Gorczyca, W., Gong, J. & Darzynkiewicz, Z. *Cancer Res.* **53**, 1945–1951 (1993).
8. Di Leonardo, A. et al. *Genes Dev.* **8**, 2540–2551 (1994).
9. Gao, X. et al. *Oncogene* **11**, 1395–1398 (1995).
10. Shibata, D. et al. *Nature Genet.* **6**, 273–281 (1994).
11. El-Deiry, W. A. et al. *Cancer Res.* **55**, 2910–2919 (1995).
12. Macleod, K. F. et al. *Genes Dev.* **8**, 935–944 (1995).
13. Kung, A. L., Sherwood, S. W. & Schimke, R. T. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9553–9557 (1990).
14. Kung, A. L., Zetterberg, A., Sherwood, S. W. & Schimke, R. T. *Cancer Res.* **50**, 7307–7317 (1990).
15. Heichman, K. A. & Roberts, J. M. *Cell* **79**, 557–562 (1994).

16. Li, J. J. & Deshaies, R. J. *Cell* **74**, 223–226 (1993).
17. Chen, J., Jackson, P. K., Kirschner, M. W. & Dutta, A. *Nature* **374**, 386–388 (1995).
18. Luo, Y., Hurwitz, J. & Massague, J. *Nature* **375**, 159–161 (1995).
19. Graeber, T. G. et al. *Nature* **379**, 88–91 (1996).
20. Cross, S. M. et al. *Science* **267**, 1353–1356 (1995).
21. Carder, P. et al. *Oncogene* **8**, 1397–1401 (1993).
22. Harvey, M. et al. *Oncogene* **8**, 2457–2467 (1993).
23. Metz, T., Harris, A. W. & Adams, J. M. *Cell* **82**, 29–36 (1995).
24. Morrow, C. S. & Cowan, K. H. in *Cancer Medicine* (eds Holland, J. F. et al.), 618–631 (Lea & Febiger, Philadelphia, 1993).
25. Ried, T. et al. *Genes, Chrom. Cancer* **4**, 69–74 (1992).
26. Lengauer, C. et al. *Genet. Anal. Tech. Applic.* **11**, 140–147 (1994).
27. Lichter, P. & Cremer, T. *Human Cytogenetics: A Practical Approach* 157–192 (IRL University Press, Oxford, 1992).
28. Giberson, R. T. & Demaree, R. S. *Microsc. Res. Tech.* **32**, 246–254 (1995).

ACKNOWLEDGEMENTS. We thank D. Vindich and D. Coffey for use of digital fluorescence equipment, J. Flook for technical assistance with flow cytometry, and M. Delannoy for help with electron microscopy. This work was supported by the Clayton Fund and NIH grants. C.L. is a recipient of an Erwin-Schrödinger-Stipendium (Fonds zur Förderung der wissenschaftlichen Forschung). B.V. is an Investigator of the Howard Hughes Medical Institute.

CORRESPONDENCE and requests for materials should be addressed to B.V.